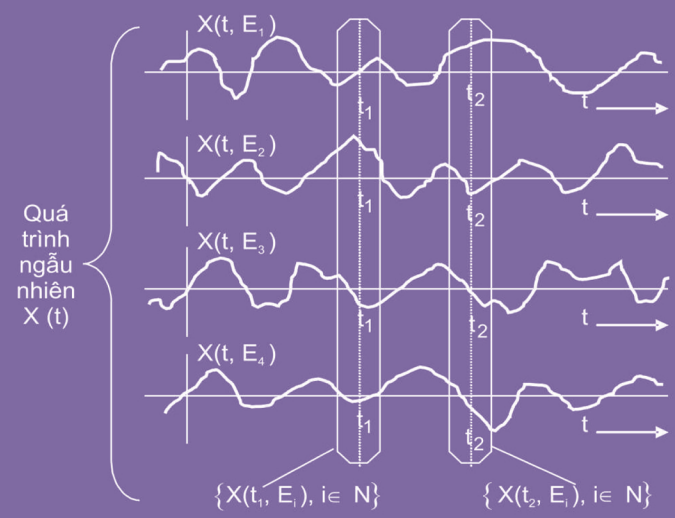
HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
PGS.TS. LÊ BÁ LONG

GIÁO TRÌNH

Xác suất và Thống kê



PUBLISHER
NHÀ XUẤT BẢN THÔNG TIN VÀ TRUYỀN THÔNG

GIÁO TRÌNH

Xác suất và Thống kê

Chịu trách nhiệm xuất bản
Giám đốc: TRẦN CHÍ ĐẠT

Biên tập: NGÔ MỸ HẠNH
NGUYỄN LONG BIÊN
Trình bày sách: NGUYỄN ĐỨC ANH
Thiết kế bìa: TRẦN HỒNG MINH
Kỹ thuật sách điện tử: NGUYỄN TIẾN PHÁT - BÙI HỮU LỘ

(Giáo trình này được ban hành kèm theo Quyết định số 51/QĐ-ĐT&KHCN ngày 20 tháng 02 năm 2009 của Giám đốc Học viện Công nghệ Bưu chính Viễn thông)

Số đăng ký kế hoạch xuất bản: 1592-2016/CXBIPH/28-66/TTTT.

Số quyết định xuất bản: 294/QĐ-NXB TTTT ngày 30 tháng 6 năm 2016.

Mã ISBN: 978-604-80-1851-1

Nộp lưu chiểu tháng 7 năm 2016.

LỜI NÓI ĐẦU

Lý thuyết xác suất và thống kê là một bộ phận của toán học nghiên cứu các hiện tượng ngẫu nhiên và ứng dụng chúng vào thực tế. Có thể coi hiện tượng ngẫu nhiên là hiện tượng không thể biết trước nó xảy ra hay không xảy ra, khi thực hiện một lần quan sát. Tuy nhiên, nếu tiến hành quan sát khá nhiều lần một hiện tượng ngẫu nhiên trong các phép thử như nhau, ta có thể rút ra được những kết luận khoa học về hiện tượng này. Lý thuyết xác suất cũng là cơ sở để nghiên cứu Thống kê - môn học nghiên cứu các phương pháp thu thập thông tin chọn mẫu, xử lý thông tin, nhằm rút ra các kết luận hoặc quyết định cần thiết.

Ngày nay, với sự hỗ trợ tích cực của máy tính điện tử và công nghệ thông tin, lý thuyết xác suất và thống kê ngày càng được ứng dụng rộng rãi và hiệu quả trong nhiều lĩnh vực của tự nhiên và xã hội. Để phục vụ nhu cầu học tập cho sinh viên, Học viện Công nghệ Bưu chính Viễn thông phối hợp với Nhà xuất bản Thông tin và Truyền thông xuất bản ***"Giáo trình Xác suất và thống kê"*** do PGS.TS. Lê Bá Long biên soạn.

Giáo trình được biên soạn cho Hệ đại học chuyên ngành Điện tử viễn thông và Công nghệ thông tin, dựa trên nền chương trình khung của Bộ Giáo dục - Đào tạo và theo đề cương chương trình của Học viện Công nghệ Bưu chính Viễn thông phê duyệt năm 2007, cùng với kinh nghiệm giảng dạy nhiều năm của tác giả theo định hướng ứng dụng cho các ngành kỹ thuật.

Nội dung giáo trình gồm 6 chương, tương ứng với 4 đơn vị học trình:

Chương 1: Các khái niệm cơ bản về xác suất

Chương 2: Biến ngẫu nhiên và các đặc trưng của chúng

Chương 3: Véc tơ ngẫu nhiên và các đặc trưng cơ bản của chúng

Chương 4: Luật số lớn và định lý giới hạn

Chương 5: Thống kê toán học

Chương 6: Quá trình ngẫu nhiên và chuỗi Markov.

Điều kiện tiên quyết cho môn học xác suất và thống kê là hai môn toán cao cấp đại số và giải tích trong chương trình toán đại cương.

Một số giáo trình trình bày lý thuyết xác suất trên cơ sở hệ tiên đề Kolmogorov, trong đó xác suất là một độ đo σ - cộng tính các định trong σ-đại số các biến cố. Với cách xây dựng này, các khái niệm và định lý được trình bày ngắn gọn hơn, mạch lạc hơn, tổng quát hơn. Tuy nhiên vì sự hạn chế về thời gian của chương trình toán dành cho ngành kỹ thuật nên tác giả lựa chọn cách tiếp cận lý thuyết xác suất theo hướng ứng dụng mà không theo hướng hệ tiên đề. Một số khái niệm được hình thành từ các ví dụ cụ thể, nhiều kết quả và định lý chỉ được phát biểu và minh họa chứ không có điều kiện để chứng minh chi tiết.

Giáo trình được trình bày phù hợp với người tự học. Trước khi nghiên cứu nội dung chi tiết, bạn đọc nên xem phần giới thiệu của mỗi chương để thấy được mục đích, ý nghĩa, yêu cầu chính của chương đó. Tại mỗi chương, mỗi nội dung bạn đọc có thể tự đọc và hiểu được cặn kẽ thông qua cách diễn đạt và chỉ dẫn rõ ràng. Đặc biệt bạn đọc nên chú ý đến các nhận xét để hiểu sâu hơn hoặc mở rộng tổng quát hơn các kết quả và hướng ứng dụng vào thực tế. Trong giáo trình tác giả đưa vào nhiều ví dụ minh họa trực tiếp cho các khái niệm, định lý hoặc các thuật toán, vì vậy sẽ giúp người đọc dễ dàng hơn khi tiếp thu bài học. Cuối mỗi chương đều có các câu hỏi luyện tập và bài tập. Bộ câu hỏi này bao trùm toàn bộ nội dung trong chương: có những câu kiểm tra trực tiếp các kiến thức vừa được học nhưng cũng có những câu đòi hỏi bạn đọc phải vận dụng một cách tổng hợp và sáng tạo các kiến thức để giải quyết. Vì vậy việc giải các bài tập này giúp bạn đọc nắm chắc hơn lý thuyết và tự kiểm tra được mức độ tiếp thu lý thuyết của mình.

Giáo trình là tài liệu học tập và tham khảo hữu ích cho sinh viên tất cả các trường đại học và cao đẳng kỹ thuật trong cả nước. Trong quá trình biên soạn khó tránh khỏi thiếu sót, Học viện rất mong nhận được ý kiến góp ý của các bạn đồng nghiệp và bạn đọc gần xa.

Trân trọng cảm ơn!

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Chương 1

CÁC KHÁI NIỆM CƠ BẢN VỀ XÁC SUẤT

GIỚI THIỆU

Các hiện tượng trong tự nhiên hay xã hội xảy ra một cách ngẫu nhiên (không biết trước kết quả) hoặc tất định (biết trước kết quả sẽ xảy ra). Chẳng hạn một vật nặng được thả từ trên cao chắc chắn sẽ rơi xuống đất, trong điều kiện bình thường nước sôi ở 100^0C... Đó là những hiện tượng diễn ra có tính quy luật, tất định. Trái lại khi tung đồng xu ta không biết mặt sấp hay mặt ngửa sẽ xuất hiện. Ta không thể biết trước có bao nhiêu cuộc gọi đến tổng đài, có bao nhiêu khách hàng đến điểm phục vụ trong khoảng thời gian nào đó. Ta không thể xác định trước chỉ số chứng khoán trên thị trường chứng khoán… Đó là những hiện tượng ngẫu nhiên. Tuy nhiên, nếu tiến hành quan sát khá nhiều lần một hiện tượng ngẫu nhiên trong những hoàn cảnh như nhau, thì trong nhiều trường hợp ta có thể rút ra những kết luận có tính quy luật về những hiện tượng này. Lý thuyết xác suất nghiên cứu các qui luật của các hiện tượng ngẫu nhiên. Việc nắm bắt các quy luật này sẽ cho phép dự báo các hiện tượng ngẫu nhiên đó sẽ xảy ra như thế nào. Chính vì vậy các phương pháp của lý thuyết xác suất được ứng dụng rộng rãi trong việc giải quyết các bài toán thuộc nhiều lĩnh vực khác nhau của khoa học tự nhiên, kỹ thuật và kinh tế-xã hội.

Nội dung chương này trình bày một cách có hệ thống các khái niệm và các kết quả chính về lý thuyết xác suất:

- Các khái niệm phép thử, biến cố.

- Quan hệ giữa các biến cố.

- Các định nghĩa về xác suất: định nghĩa xác suất theo cổ điển, theo thống kê và theo hình học.

- Các tính chất của xác suất: công thức cộng và xác suất của biến cố đối.

- Xác suất có điều kiện, công thức nhân trong trường hợp không độc lập. Công thức xác suất đầy đủ và định lý Bayes.

- Dãy phép thử Bernoulli và xác suất nhị thức.

Khi đã nắm vững các kiến thức về đại số tập hợp (một trường hợp cụ thể của đại số Boole) như hợp, giao tập hợp, tập con, phần bù của một tập con… bạn đọc sẽ dễ dàng trong việc tiếp thu, biểu diễn hoặc mô tả các biến cố.

Để tính xác suất các biến cố theo phương pháp cổ điển đòi hỏi phải tính số các trường hợp thuận lợi đối với biến cố và số các trường hợp đồng khả năng. Vì vậy bạn đọc cần nắm vững các phương pháp đếm - giải tích tổ hợp (đã được học ở lớp 12 và trong chương 1 của môn đại số). Tuy nhiên để thuận lợi cho người học chúng tôi sẽ nhắc lại các kết quả chính về phương pháp đếm trong mục 1.2.2.

Một trong những khó khăn của bài toán xác suất là xác định được biến cố và sử dụng đúng các công thức thích hợp. Bằng cách tham khảo các ví dụ và giải nhiều bài tập bạn đọc sẽ rèn luyện tốt kỹ năng này.

1.1 PHÉP THỬ VÀ BIẾN CỐ

1.1.1 Phép thử (Experiment)

Trong thực tế ta thường gặp nhiều thí nghiệm, mà khi quan sát các kết quả của nó không thể dự báo trước được. Ta gọi chúng là các phép thử ngẫu nhiên.

Với phép thử gieo con xúc xắc, tuy không biết kết quả sẽ xảy ra như thế nào, nhưng ta có thể liệt kê được hoặc biểu diễn tất cả các kết quả của phép thử này; đó là sự xuất hiện mặt có số chấm 1, 2, 3, 4, 5, 6.

Ta xem các kết quả này là các *biến cố sơ cấp*. Tập hợp tất cả các biến cố sơ cấp của phép thử được gọi là *không gian mẫu*, ký hiệu Ω. Vậy $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Ví dụ 1.1:

- Phép thử tung đồng xu có không gian mẫu là $\Omega = \{S, N\}$.

- Phép thử tung đồng thời 2 đồng xu có không gian mẫu là:

$$\Omega = \{(S,S),(S,N),(N,S),(N,N)\}.$$

Chú ý rằng bản chất của các biến cố sơ cấp không có vai trò đặc biệt gì trong lý thuyết xác suất. Chẳng hạn có thể mã hóa các kết quả và xem không gian mẫu của phép thử tung đồng xu là $\Omega = \{0, 1\}$, trong đó 0 là biến cố sơ cấp chỉ mặt sấp xuất hiện và 1 để chỉ mặt ngửa xuất hiện.

1.1.2 Biến cố (Event)

Với phép thử $\mathscr{C}$ ta thường xét các biến cố (còn gọi là sự kiện) mà việc xảy ra hay không xảy ra biến cố đó hoàn toàn được xác định bởi kết quả của $\mathscr{C}$.

Mỗi kết quả ω của $\mathscr{C}$ được gọi là kết quả thuận lợi cho biến cố A nếu A xảy ra khi kết quả của phép thử $\mathscr{C}$ là ω.

Ví dụ 1.2: Nếu gọi A là biến cố "số chấm xuất hiện là chẵn" trong phép thử tung xúc sắc ở ví dụ 1.1 thì A có các kết quả thuận lợi là 2, 4, 6.

Tung hai đồng xu, biến cố xuất hiện một mặt sấp một mặt ngửa (xin âm dương) có các kết quả thuận lợi là (S,N); (N,S).

Như vậy mỗi biến cố A có thể đồng nhất với một tập con của không gian mẫu Ω bao gồm các kết quả thuận lợi đối với A.

Mỗi biến cố chỉ có thể xảy ra khi một phép thử được thực hiện, nghĩa là gắn với không gian mẫu nào đó.

Có hai biến cố đặc biệt sau:

- *Biến cố chắc chắn* là biến cố luôn luôn xảy ra khi thực hiện phép thử. Không gian mẫu Ω là một biến cố chắc chắn.

- *Biến cố không thể* là biến cố nhất định không xảy ra khi thực hiện phép thử. Biến cố không thể được ký hiệu $\varnothing$.

Tung một con xúc xắc, biến cố xuất hiện mặt có số chấm nhỏ hơn hay bằng 6 là biến chắc chắn, biến cố xuất hiện mặt có 7 chấm là biến cố không thể.

1.1.3 Quan hệ giữa các biến cố

Một cách tương ứng với các phép toán của tập hợp, trong lý thuyết xác suất người ta xét các quan hệ sau đây cho các biến cố trong cùng một phép thử.

a) Quan hệ kéo theo

Biến cố A kéo theo biến cố B, ký hiệu $A \subset B$, nếu khi A xảy ra thì B xảy ra.

Nếu $A \subset B$ và $B \subset A$ thì ta nói hai biến cố A, B trùng nhau, ký hiệu $A = B$.

b) Quan hệ biến cố đối

Với mỗi biến cố A, luôn có biến cố gọi là biến cố đối của A, ký hiệu $\overline{A}$ và được xác định như sau: A xảy ra khi và chỉ khi $\overline{A}$ không xảy ra.

Ví dụ 1.3: Bắn một phát đạn vào bia. Gọi A là biến cố "bắn trúng bia". Biến cố đối của A là $\overline{A}$ "bắn trượt bia".

c) Tổng của hai biến cố

Tổng của hai biến cố A, B là biến cố được ký hiệu $A \cup B$. Biến cố $A \cup B$ xảy ra khi và chỉ khi có ít nhất A hoặc B xảy ra.

Tổng của một dãy các biến cố $\{A_1, A_2, ..., A_n\}$ là biến cố $\bigcup_{i=1}^{n} A_i$. Biến cố này xảy ra khi có ít nhất một trong các biến cố A_i xảy ra ($i = 1, ..., n$).

Ví dụ 1.4: Một mạng điện gồm hai bóng đèn mắc nối tiếp. Gọi A_1 là biến cố "bóng đèn thứ nhất bị cháy", A_2 là biến cố "bóng đèn thứ hai bị cháy". Gọi A là biến cố "mạng mất điện". Ta thấy rằng

mạng bị mất điện khi ít nhất một trong hai bóng bị cháy. Vậy $A = A_1 \cup A_2$.

d) Tích của hai biến cố

Tích của hai biến cố A, B là biến cố được ký hiệu AB. Biến cố AB xảy ra khi cả hai biến cố A, B đồng thời cùng xảy ra.

Tích của một dãy các biến cố $\{A_1, A_2, ..., A_n\}$ là biến cố $\prod_{i=1}^{n} A_i$. Biến cố này xảy ra khi tất cả các biến cố A_i cùng xảy ra ($i = 1, ..., n$).

Ví dụ 1.5: Một mạng điện gồm hai bóng đèn mắc song song. Gọi A_1 là biến cố "bóng đèn thứ nhất bị cháy", A_2 là biến cố "bóng đèn thứ hai bị cháy". Gọi A là biến cố "mạng mất điện".

Ta thấy rằng mạng bị mất điện khi cả hai bóng bị cháy. Vậy $A = A_1 A_2$.

e) Biến cố xung khắc

Hai biến số A, B gọi là xung khắc nếu biến cố tích AB là biến cố không thể. Nghĩa là hai biến cố này không thể đồng thời cùng xảy ra.

Ví dụ 1.6: Một bình có 3 loại cầu: cầu màu trắng, màu đỏ và màu xanh. Lấy ngẫu nhiên 1 cầu từ bình. Gọi A_t, $A_đ$, A_x lần lượt là biến cố quả cầu rút được là cầu trắng, đỏ, xanh. Các biến cố này xung khắc từng đôi một, vì mỗi quả cầu chỉ có 1 màu.

Chú ý rằng các biến cố với phép toán tổng, tích và lấy biến cố đối tạo thành đại số Boole do đó các phép toán được định nghĩa ở trên có các tính chất như các phép toán hợp, giao, lấy phần bù đối với các tập con của không gian mẫu.

f) Hệ đầy đủ các biến cố

Dãy các biến cố $A_1, A_2, ..., A_n$ được gọi là một hệ đầy đủ các biến cố nếu:

(i) Xung khắc từng đôi một, nghĩa là $A_iA_j = \varnothing$ với mọi $i \neq j$; $i = 1,...,n$; $j = 1,...,n$

(ii) Tổng của chúng là biến cố chắc chắn, nghĩa là $\bigcup_{i=1}^{n} A_i = \Omega$.

Đặc biệt với mọi biến cố A, hệ $\{A, \overline{A}\}$ là hệ đầy đủ.

Ví dụ 1.7: Một nhà máy có ba phân xưởng sản xuất ra cùng một loại sản phẩm. Giả sử rằng mỗi sản phẩm của nhà máy chỉ do một trong ba phân xưởng này sản xuất. Chọn ngẫu nhiên một sản phẩm, gọi A_1, A_2, A_3 lần lượt là biến cố sản phẩm được chọn do phân xưởng thứ nhất, thứ hai, thứ ba sản xuất. Khi đó hệ ba biến cố $\{A_1, A_2, A_3\}$ là hệ đầy đủ.

g) Tính độc lập của các biến cố

Hai biến cố A và B được gọi là độc lập với nhau nếu việc xảy ra hay không xảy ra biến cố này không ảnh hưởng tới việc xảy ra hay không xảy ra biến cố kia.

Tổng quát các biến cố $A_1, A_2, ..., A_n$ được gọi là độc lập nếu việc xảy ra hay không xảy ra của một nhóm bất kỳ k biến cố, trong đó $1 \leq k \leq n$, không làm ảnh hưởng tới việc xảy ra hay không xảy ra của một nhóm nào đó các biến cố còn lại.

Định lý 1.1: Nếu A, B độc lập thì các cặp biến cố: $A, \overline{B}$; $\overline{A}, B$; $\overline{A}, \overline{B}$ cũng độc lập.

Ví dụ 1.8: Ba xạ thủ $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ mỗi người bắn một viên đạn vào mục tiêu. Gọi A, B, C lần lượt là biến cố $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ bắn trúng mục tiêu.

a. Hãy mô tả các biến cố: ABC, $\overline{A}\,\overline{B}\,\overline{C}$, $A \cup B \cup C$.

b. Biểu diễn các biến cố sau theo A, B, C:

- D: Có ít nhất 2 xạ thủ bắn trúng.

- E : Có nhiều nhất 1 xạ thủ bắn trúng.
- F : Chỉ có xạ thủ $\mathscr{C}$ bắn trúng.
- G : Chỉ có 1 xạ thủ bắn trúng.

c. Các biến cố A, B, C có xung khắc, có độc lập không?

Giải:

a. ABC : cả 3 đều bắn trúng. $\overline{A}\,\overline{B}\,\overline{C}$: cả 3 đều bắn trượt. $A \cup B \cup C$: có ít nhất 1 người bắn trúng.

b. $D = AB \cup BC \cup CA$.

Có nhiều nhất một xạ thủ bắn trúng có nghĩa là có ít nhất hai xạ thủ bắn trượt, vậy

$$E = \overline{A}\,\overline{B} \cup \overline{B}\,\overline{C} \cup \overline{C}\,\overline{A}$$

$$F = \overline{A}\,\overline{B}\,C$$

$$G = A\overline{B}\,\overline{C} \cup \overline{A}\,B\overline{C} \cup \overline{A}\,\overline{B}\,C$$

c. Ba biến cố A, B, C độc lập vì biến cố bắn trúng mục tiêu của mỗi xạ thủ là độc lập nhau.

Ba biến cố *A, B, C* không xung khắc vì có thể cùng bắn trúng mục tiêu.

1.2 ĐỊNH NGHĨA XÁC SUẤT VÀ CÁC TÍNH CHẤT

Một biến cố ngẫu nhiên xảy ra hay không trong kết quả của một phép thử là điều không thể biết hoặc đoán trước được. Tuy nhiên bằng những cách khác nhau ta có thể định lượng khả năng xuất hiện của biến cố, đó là xác suất xuất hiện của biến cố.

Xác suất của một biến cố là một con số đặc trưng khả năng khách quan xuất hiện biến cố đó khi thực hiện phép thử.

Xác suất của biến cố A ký hiệu $P(A)$. Trường hợp biến cố chỉ gồm một biến cố sơ cấp $\{a\}$ ta ký hiệu $P(a)$ thay cho $P(\{a\})$.

Dựa vào bản chất của phép thử (đồng khả năng) ta có thể suy luận về khả năng xuất hiện của biến cố, với cách tiếp cận này ta có định nghĩa xác suất theo phương pháp cổ điển.

Khi thực hiện nhiều lần lặp lại độc lập một phép thử ta có thể tính được tần suất xuất hiện của một biến cố nào đó. Tần suất thể hiện khả năng xuất hiện của biến cố, với cách tiếp cận này ta có định nghĩa xác suất theo thống kê.

1.2.1 Định nghĩa cổ điển về xác suất

Định nghĩa 1.1: *Giả sử phép thử $\mathscr{C}$ thoả mãn hai điều kiện sau:*

(i) Không gian mẫu có một số hữu hạn phần tử.

(ii) Các kết quả xảy ra đồng khả năng.

Khi đó ta định nghĩa xác suất của biến cố A là

$$P(A)=\frac{\text{Số trường hợp thuận lợi đối với } A}{\text{Số trường hợp có thể}} \tag{1.1a}$$

Nếu xem biến cố A như là tập con của không gian mẫu Ω thì

$$P(A)=\frac{\text{Số phần tử của } A}{\text{Số phần tử của } \Omega}=\frac{|A|}{|\Omega|} \tag{1.1b}$$

Ví dụ 1.9: Biến cố A xuất hiện mặt chẵn trong phép thử gieo con xúc sắc ở ví dụ 1.1 có 3 trường hợp thuận lợi ($|A|=3$) và 6 trường hợp có thể ($|\Omega|=6$). Vậy $P(A)=\frac{3}{6}=\frac{1}{2}$.

Biến cố xuất hiện một mặt sấp và một mặt ngửa khi gieo đồng thời hai đồng xu có 2 kết quả thuận lợi và 4 kết quả đồng khả năng có thể, vậy xác suất xuất hiện của biến cố đó là $\frac{1}{2}$.

Để tính xác suất cổ điển ta sử dụng phương pháp đếm của giải tích tổ hợp.

Ví dụ 1.10: Xét phép thử gieo liên tiếp 2 lần con xúc sắc 4 mặt (hình tứ diện). Tính xác suất của các biến cố sau:

a) Tổng số chấm xuất hiện là chẵn (biến cố A).

b) Số chấm xuất hiện của hai con xúc sắc bằng nhau (biến cố B).

c) Số chấm của xúc sắc thứ nhất lớn hơn xúc sắc thứ hai (biến cố C).

d) Ít nhất một xúc sắc xuất hiện mặt 4 chấm (biến cố D).

Giải: Có thể biểu diễn không gian mẫu của phép thử và các biến cố tương ứng dưới dạng biểu đồ (xem hình 1.1):

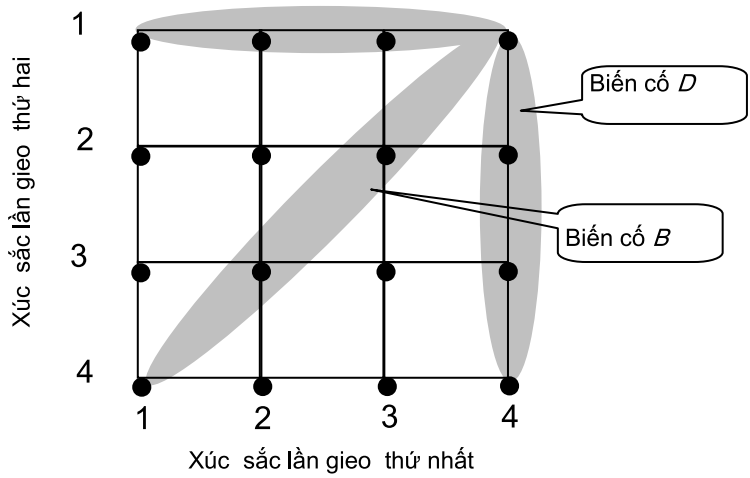
Từ biểu đồ ta thấy có 16 trường hợp đồng khả năng có thể và các xác suất tương ứng:

a) $P(A) = \frac{8}{16} = \frac{1}{2}$.

b) $P(B) = \frac{4}{16} = \frac{1}{4}$.

c) Số các trường hợp thuận lợi của biến cố C là số các điểm ở dưới đường chéo thứ hai trong hình 1.1, vậy $P(C) = \frac{6}{16} = \frac{3}{8}$.

d) $P(D) = \frac{7}{16}$.

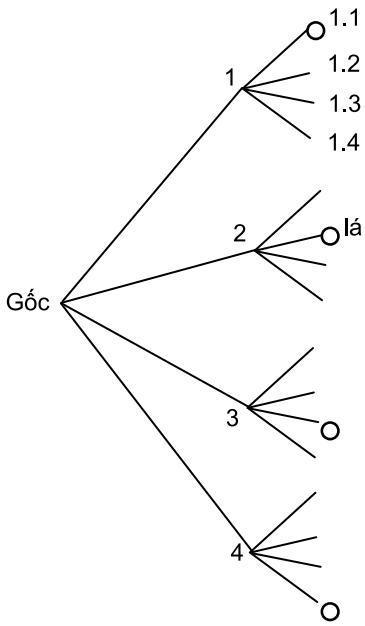


Hình 1.1: Phép thử gieo 2 xúc sắc 4 mặt

Ví dụ 1.11: *Sơ đồ cây*

Nhiều phép thử có tính chất nối tiếp lập thành dãy, chẳng hạn phép thử tung liên tiếp đồng xu ba lần, quan sát chỉ số chứng khoán trong năm ngày liên tiếp, hoặc tám ký số liên tiếp nhận được của một bộ nhận thông tin,... Trong trường hợp này ta có thể biểu diễn không gian mẫu và các biến cố tương ứng dưới dạng sơ đồ cây.

Không gian mẫu và biến cố B của ví dụ 1.10 được biểu diễn dạng sơ đồ cây như sau:



Hình 1.2: Sơ đồ cây của phép thử gieo 2 xúc sắc 4 mặt

1.2.2 Các qui tắc đếm

a) Qui tắc cộng

Nếu có m_1 cách chọn loại đối tượng x_1, m_2 cách chọn loại đối tượng x_2,..., m_n cách chọn loại đối tượng x_n. Các cách chọn đối tượng x_i không trùng với cách chọn x_j nếu $i \neq j$ thì có $m_1 + m_2 + \cdots + m_n$ cách chọn một trong các đối tượng đã cho.

b) Qui tắc nhân

Giả sử công việc H gồm nhiều công đoạn liên tiếp H_1, H_2,..., H_k và mỗi công đoạn H_i có n_i cách thực hiện thì có tất cả $n_1 \times n_2 \times \cdots \times n_k$ cách thực hiện công việc H.

c) Hoán vị

Mỗi phép đổi chỗ của n phần tử hoặc xếp n phần tử vào n vị trí được gọi là phép hoán vị n phần tử. Sử dụng quy tắc nhân ta có thể tính được:

Có $n!$ hoán vị n phần tử.

Quy ước 0! = 1.

d) Chỉnh hợp có lặp

Chọn lần lượt k phần tử hoàn lại trong tập n phần tử ta được một chỉnh hợp lặp chập k của n phần tử. Sử dụng quy tắc nhân ta có thể tính được số các chỉnh hợp lặp chập k của n phần tử là n^k.

e) Chỉnh hợp

Chọn lần lượt k $(1 \le k \le n)$ phần tử không hoàn lại trong tập n phần tử ta được một chỉnh hợp chập k của n phần tử. Sử dụng quy tắc nhân ta có thể tính được số các chỉnh hợp chập k của n phần tử là

$$A_n^k = \frac{n!}{(n-k)!} = n \cdot (n-1) \cdots (n-k+1) \qquad (1.2)$$

f) Tổ hợp

Một tổ hợp chập k $(1 \le k \le n)$ của n phần tử là một cách chọn đồng thời k phần tử từ một tập có n phần tử. Vì vậy cũng có thể xem một tập con k phần tử của tập n phần tử là một tổ hợp chập k của n phần tử.

Hai chỉnh hợp chập k của n phần tử là khác nhau nếu thỏa mãn một trong hai điều kiện sau:

- Có ít nhất 1 phần tử của chỉnh hợp này không có trong chỉnh hợp kia.

- Các phần tử đều như nhau nhưng thứ tự khác nhau.

Do đó với mỗi tổ hợp chập k của n phần tử có $k!$ chỉnh hợp tương ứng. Mặt khác hai chỉnh hợp khác nhau ứng với hai tổ hợp khác nhau là khác nhau.

Vậy số các tổ hợp chập k của n phần tử là

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!} \tag{1.3}$$

Ví dụ 1.12: Tung một con xúc sắc (6 mặt) hai lần. Tìm xác suất để trong đó chỉ có 1 lần ra 6 chấm.

Giải: Số các trường hợp có thể là 36. Gọi A là biến cố "trong 2 lần tung con xúc sắc chỉ có 1 lần được mặt 6". Nếu lần thứ nhất ra mặt 6 thì lần thứ hai chỉ có thể ra các mặt từ 1 đến 5, nghĩa là có 5 trường hợp. Tương tự cũng có 5 trường hợp ứng với xuất hiện mặt 6 ở lần tung thứ hai. Áp dụng quy tắc cộng ta suy ra xác suất để chỉ có một lần ra mặt 6 khi tung xúc sắc 2 lần là $\frac{10}{36}$.

Ví dụ 1.13: Bố trí một cách ngẫu nhiên n người ngồi xung quanh một ban tròn ($n \geq 3$), trong đó có hai người là anh em. Tìm xác suất để hai anh em ngồi cạnh nhau.

Giải: Chúng ta đánh số ghế ngồi từ 1 đến n và coi 2 cách ngồi là khác nhau nếu có ít nhất 1 chỗ lần lượt có 2 người ngồi khác nhau.

Số trường hợp có thể là số hoán vị n phần tử: $n!$

Ta xếp người anh ngồi tùy ý vào 1 trong n chỗ (có n cách); người em ngồi vào 1 trong 2 chỗ cạnh người anh (có 2 cách); $n-2$ người còn lại còn lại ngồi tùy ý vào $n-2$ chỗ còn lại (có $(n-2)!$ cách). Vậy số các trường hợp thuận lợi là $(n)(2)\big((n-2)!\big)$.

Xác suất cần tìm $P = \dfrac{(n)(2)\big((n-2)!\big)}{n!} = \dfrac{2}{n-1}$.

Ví dụ 1.14: Một người gọi điện thoại quên mất hai số cuối của số điện thoại và chỉ nhớ được rằng chúng khác nhau. Tìm xác suất để quay ngẫu nhiên một lần được đúng số cần gọi.

Giải: Gọi A là biến cố "quay ngẫu nhiên một lần được đúng số cần gọi". Số các trường hợp có thể là số các cặp hai chữ số khác nhau từ 10 chữ số từ 0 đến 9. Số các cặp hai chữ số này bằng số các chỉnh hợp chập 2 của 10. Vậy số các trường hợp có thể là $A_{10}^2 = 10 \cdot 9 = 90$.

Chỉ có 1 trường hợp thuận lợi đối với A. Do đó $P(A) = \dfrac{1}{90}$.

Ví dụ 1.15: Cho các từ mã 6 bit được tạo từ các chuỗi các bit 0 và bit 1 đồng khả năng. Hãy tìm xác suất của các từ có chứa k bit 1, với các trường hợp $k = 0,...,6$.

Giải: Số trường hợp có thể $|\Omega| = 2^6$. Đặt A_k là biến cố "từ mã có chứa k bit 1". Có thể xem mỗi từ mã có chứa k bit 1 là một tổ hợp chập k của 6 phần tử, vậy số trường hợp thuận lợi đối với A_k là số các tổ hợp chập k của 6 phần tử. Do đó $|A_k| = C_6^k = \dfrac{6!}{k!(6-k)!}$

Vậy xác suất của các biến cố tương ứng:

$$P(A_k) = \frac{6!}{k!(6-k)!2^6},\ k = 0,...,\ 6.$$

Ví dụ 1.16: Một công ty cần tuyển 2 nhân viên. Có 6 người nộp đơn trong đó có 4 nữ và 2 nam. Giả sử khả năng trúng tuyển của cả 6 người là như nhau. Tính xác suất các biến cố:

a) Hai người trúng tuyển là nam

b) Hai người trúng tuyển là nữ

c) Có ít nhất 1 nữ trúng tuyển.

Giải: Số trường hợp có thể $|\Omega| = C_6^2 = 15$.

a) Chỉ có 1 trường hợp cả 2 nam đều trúng tuyển do đó xác suất tương ứng là $P = 1/15$.

b) Có $C_4^2 = 6$ cách chọn 2 nữ trong 4 nữ, vậy xác suất tương ứng $P = 6/15$.

c) Trong 15 trường hợp có thể chỉ có 1 trường hợp cả 2 nam được chọn, vậy có 14 trường hợp ít nhất 1 nữ được chọn. Do đó xác suất tương ứng $P = 14/15$.

Có thể tính số trường hợp thuận lợi của biến cố "có ít nhất 1 nữ được chọn" như sau:

- Có $C_4^2 = 6$ cách chọn 2 nữ trong 4 nữ.

- Có $C_4^1 = 4$ cách chọn 1 nữ trong 4 nữ và có $C_2^1 = 2$ cách chọn 1 nam trong 2 nam.

Vậy có $6 + 4 \times 2 = 14$ trường hợp thuận lợi của biến cố "có ít nhất 1 nữ được chọn".

1.2.3 Định nghĩa xác suất theo thống kê

Định nghĩa xác suất theo cổ điển trực quan, dễ hiểu. Tuy nhiên khi số các kết quả có thể vô hạn hoặc không đồng khả năng thì cách tính xác suất cổ điển không áp dụng được. Trong trường hợp này có thể sử dụng phương pháp thống kê để xác định xác suất theo định nghĩa sau:

Định nghĩa 1.2: *Giả sử phép thử $\mathscr{C}$ có thể được thực hiện lặp lại nhiều lần độc lập trong những điều kiện giống hệt nhau. Nếu trong n lần thực hiện phép thử $\mathscr{C}$, biến cố A xuất hiện $k_n(A)$ lần thì tỉ số*

$$f_n(A) = \frac{k_n(A)}{n} \tag{1.4}$$

được gọi là tần suất xuất hiện của biến cố A trong n phép thử.

Người ta chứng minh được (định lý luật số lớn) khi n tăng lên vô hạn thì $f_n(A)$ tiến đến một giới hạn xác định.

Ta định nghĩa giới hạn này là xác suất của biến cố A, ký hiệu P(A)

$$P(A) = \lim_{n\to\infty} f_n(A) \qquad (1.5)$$

Trên thực tế các tần suất $f_n(A)$ xấp xỉ nhau khi n đủ lớn. Vì vậy $P(A)$ được chọn bằng giá trị xấp xỉ này.

Ví dụ 1.17: Một công ty bảo hiểm muốn xác định xác suất để một người Mỹ 25 tuổi sẽ bị chết trong năm tới. Để làm điều đó người ta theo dõi 100.000 thanh niên và thấy rằng có 798 người bị chết trong vòng 1 năm sau đó. Vậy xác suất cần tìm xấp xỉ bằng 0,008.

Ví dụ 1.18: Thống kê cho thấy tần suất sinh con trai xấp xỉ 0,513. Vậy xác suất để bé trai ra đời lớn hơn bé gái.

Nhận xét 1.1: Định nghĩa xác suất theo thống kê khắc phục được hạn chế của định nghĩa cổ điển, nó hoàn toàn dựa trên các thí nghiệm quan sát thực tế để tìm xác suất của biến cố. Tuy nhiên định nghĩa thống kê về xác suất cũng chỉ áp dụng cho các phép thử mà có thể lặp lại được nhiều lần một cách độc lập trong những điều kiện giống hệt nhau. Ngoài ra để xác định một cách tương đối chính xác giá trị của xác suất thì cần tiến hành một số n đủ lớn lần các phép thử, mà việc này đôi khi không thể làm được vì hạn chế về thời gian và kinh phí.

Ngày nay với sự trợ giúp của công nghệ thông tin, người ta có thể mô phỏng các phép thử ngẫu nhiên mà không cần thực hiện các phép thử trong thực tế. Điều này cho phép tính xác suất theo phương pháp thống kê thuận tiện hơn.

1.2.4 Định nghĩa xác suất theo hình học

Định nghĩa 1.3: *Giả sử không gian mẫu Ω có thể biểu diễn tương ứng với một miền nào đó có diện tích (thể tích, độ dài) hữu hạn và biến cố A tương ứng với một miền con của Ω thì xác suất của biến cố A được định nghĩa:*

$$P(A) = \frac{\text{diện tích } A}{\text{diện tích } \Omega} \qquad (1.6)$$

Ví dụ 1.19: Hai người bạn X, Y hẹn gặp nhau ở một địa điểm trong khoảng thời gian từ 12h đến 13h. Mỗi người có thể đến điểm hẹn một cách ngẫu nhiên tại một thời điểm nào đó trong khoảng thời gian nói trên và họ quy ước rằng ai đến trước thì chỉ đợi người kia trong vòng 15 phút. Tính xác suất để hai người gặp nhau.

Giải: Giả sử x là thời điểm X đến điểm hẹn, y là thời điểm Y đến điểm hẹn thì:

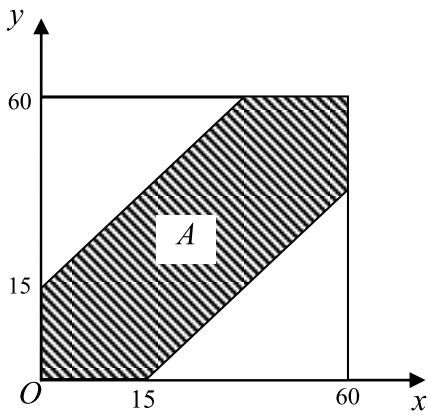
$$0 \le x \le 60\,,\ 0 \le y \le 60\,.$$

Vậy mỗi cặp thời điểm đến $(x;y)$ là một điểm của hình vuông $\Omega = [0,60]^2$ (hình 1.3).

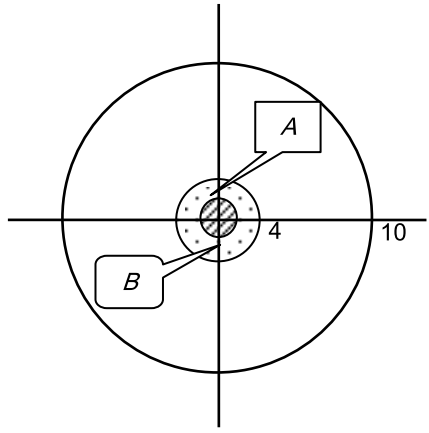
Gọi A là biến cố hai người gặp nhau, thì ta có:

$$A = \left\{ (x;y) \in \Omega \,\middle|\, |x-y| \le 15 \right\} = \left\{ (x;y) \in \Omega \,\middle|\, -15 + x \le y \le x + 15 \right\}.$$

$$\Rightarrow P(A) = \frac{\text{diện tích } A}{\text{diện tích } \Omega} = 1 - \frac{45^2}{60^2} = 1 - \frac{9}{16} = \frac{7}{16}\,.$$



Hình 1.3



Hình 1.4

Ví dụ 1.20: Xét trò chơi ném phi tiêu vào một đĩa hình tròn bán kính $10\,\text{cm}$. Nếu mũi phi tiêu cắm vào đĩa cách tâm $\le 2\,\text{cm}$ thì được giải nhất, nếu khoảng cách này ở trong khoảng $2\,\text{cm}$ đến $\le 4\,\text{cm}$ nhận

được giải thứ hai. Giả sử mũi phi tiêu luôn cắm vào trong đĩa và đồng khả năng. Tính xác suất để người chơi được giải nhất, được giải nhì.

Giải: Gọi A là biến cố người chơi nhận được giải nhất, B là biến cố người chơi nhận được giải nhì.

Có thể biểu diễn không gian mẫu Ω là hình tròn bán kính 10 (hình 1.4). Khi đó biến cố A là hình tròn cùng tâm có bán kính 2 và biến cố B là hình vành khăn bán kính đường tròn trong bằng 2 và bán kính đường tròn ngoài bằng 4. Vậy xác suất để người chơi được giải nhất, được giải nhì lần lượt là:

$$P(A) = \frac{\text{diện tích } A}{\text{diện tích } \Omega} = \frac{\pi.2^2}{\pi.10^2} = \frac{2}{50}$$

$$P(B) = \frac{\pi.(4^2 - 2^2)}{\pi.10^2} = \frac{7}{50}$$

Ta đã có ba cách tiếp cận khác nhau về xác suất một biến cố. Các định nghĩa này có các tính chất sau (mục 1.2.5).

1.2.5 Các tính chất và định lý xác suất

1.2.5.1 Các tính chất của xác suất

Các định nghĩa đã trình bày trên của xác suất thoả mãn các tính chất sau:

1. Với mọi biến cố A:

$$0 \le P(A) \le 1. \tag{1.7}$$

2. Xác suất của biến cố không thể bằng 0, xác suất của biến cố chắc chắn bằng 1.

$$P(\varnothing) = 0,\ P(\Omega) = 1 \tag{1.8}$$

1.2.5.2 Qui tắc cộng xác suất

a. Trường hợp xung khắc

Nếu A, B là hai biến cố xung khắc thì:

$$P(A \cup B) = P(A) + P(B) \tag{1.9a}$$

Tổng quát hơn, nếu $\{A_1, A_2, ..., A_n\}$ là dãy các biến cố xung khắc từng đôi một thì:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) \tag{1.9b}$$

Từ công thức (1.8) và (1.9b) ta có hệ quả: Nếu $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ thì:

$$\sum_{i=1}^{n} P(A_i) = 1 \tag{1.10}$$

b. Trường hợp tổng quát

- Nếu A, B là hai biến cố bất kỳ thì:

$$P(A \cup B) = P(A) + P(B) - P(AB) \tag{1.11a}$$

- Nếu A, B, C là ba biến cố bất kỳ thì:

$$\begin{aligned} P(A \cup B \cup C) = {} & P(A) + P(B) + P(C) - P(AB) \\ & - P(BC) - P(CA) + P(ABC) \end{aligned} \tag{1.11b}$$

- Nếu $\{A_1, A_2, ..., A_n\}$ là dãy các biến cố bất kỳ:

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) = {} & \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i A_j) \\ & + \sum_{i<j<k} P(A_i A_j A_k) - \cdots + (-1)^{n-1} P(A_1 A_2 ... A_n) \end{aligned} \tag{1.11c}$$

Ví dụ 1.21: Một lô hàng có 25% sản phẩm loại I, 55% sản phẩm loại II và 20% sản phẩm loại III. Sản phẩm được cho là đạt chất lượng nếu thuộc loại I hoặc loại II. Chọn ngẫu nhiên 1 sản phẩm tìm xác suất để sản phẩm này đạt tiêu chuẩn chất lượng.

Giải: Gọi A_1, A_2, A_3 lần lượt là biến cố sản phẩm được chọn thuộc loại I, II, III. Ba biến cố này xung khắc từng đôi một. $P(A_1) = 0{,}25$, $P(A_2) = 0{,}55$, $P(A_3) = 0{,}20$. Gọi A là biến cố sản phẩm được chọn đạt tiêu chuẩn chất lượng. Vậy $A = A_1 \cup A_2$.

$$P(A) = P(A_1) + P(A_2) = 0{,}25 + 0{,}55 = 0{,}8$$

1.2.5.3 Quy tắc tính xác suất của biến cố đối

Áp dụng công thức (1.10) cho hệ đầy đủ $\{A, \overline{A}\}$ ta được quy tắc tính xác suất biến cố đối:

Với mọi biến cố A

$$P(\overline{A}) = 1 - P(A);\ P(A) = 1 - P(\overline{A}) \tag{1.12}$$

Ví dụ 1.22: Trong phòng có n người ($n < 365$).

a) Tính xác suất có ít nhất hai người có cùng ngày sinh?

b) Tính xác suất này khi $n = 10$.

Giải: a) Gọi A là biến cố có ít nhất hai người trong phòng có cùng ngày sinh. Biến cố đối $\overline{A}$ là biến cố mọi người không trùng ngày sinh.

Vậy:

$$P(\overline{A}) = \frac{A_{365}^{n}}{365^{n}} = \frac{(365)(364)...(365-n+1)}{365^{n}},\ P(A) = 1 - P(\overline{A}).$$

b) Khi $n = 10$ thì:

$$P(\overline{A}) = \frac{A_{365}^{10}}{365^{10}} = 0,883,\ P(A) = 1 - 0,883 = 0,117.$$

Ví dụ 1.23: Gieo liên tiếp một đồng xu 3 lần.

Gọi A là biến cố lần thứ nhất ra mặt sấp. B là biến cố lần thứ hai ra mặt ngửa.

Từ sơ đồ hình 1.5 ta có:

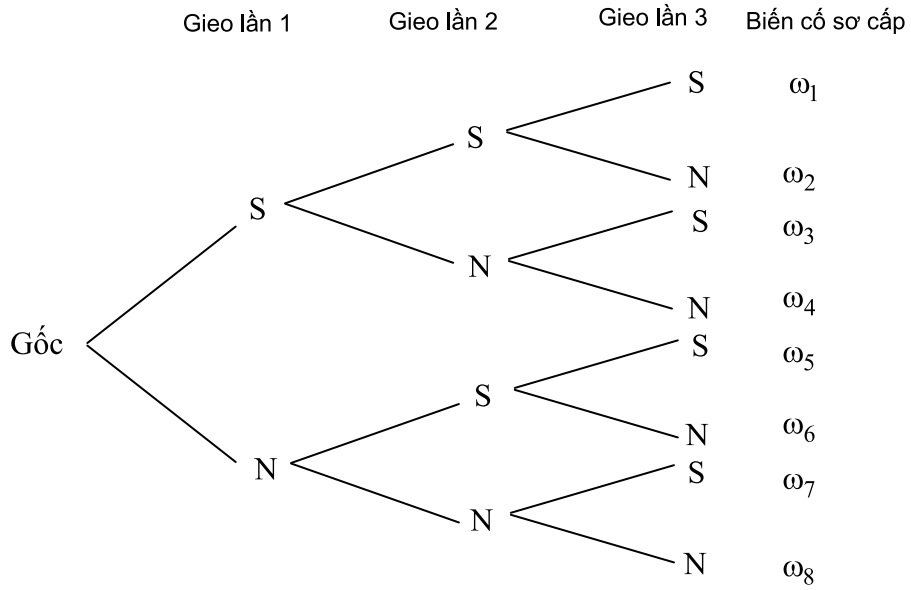
$$P(A) = P(B) = \frac{1}{2}$$

Và $AB = \{\omega_3, \omega_4\}$, do đó $P(AB) = \frac{1}{4}$. Áp dụng quy tắc cộng ta được:

$$P(A \cup B) = P(A) + P(B) - P(AB) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}.$$

Ta cũng có thể tính trực tiếp bằng cách xác định:

$A \cup B = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_7, \omega_8\}$. Vậy cũng có: $P(A \cup B) = \frac{6}{8} = \frac{3}{4}$.



Hình 1.5: Sơ đồ cây của phép thử gieo đồng xu liên tiếp 3 lần

Ví dụ 1.24: Giả sử phép thử $\mathscr{C}$ có không gian mẫu $\Omega = \{a, b, c, d\}$ với xác suất

$$P(a) = 0,2\ ;\ P(b) = 0,3\ ;\ P(c) = 0,4\ ;\ P(d) = 0,1.$$

Xét hai biến cố $A = \{a, b\}$ và $B = \{b, c, d\}$.

Tính xác suất của các biến cố $P(A)$; $P(B)$; $P(\bar{A})$; $P(A \cup B)$ và $P(AB)$.

Giải:

$$P(A) = P(a) + P(b) = 0,2 + 0,3 = 0,5$$

$$P(B) = P(b) + P(c) + P(d) = 0,3 + 0,4 + 0,1 = 0,8$$

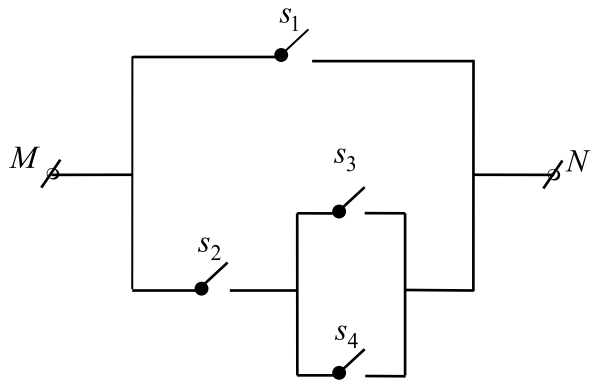
$$P(\overline{A}) = P(c) + P(d) = 0,4 + 0,1 = 0,5$$

hoặc $P(\overline{A}) = 1 - P(A) = 1 - 0,5 = 0,5$

$A \cup B = \Omega$ do đó $P(A \cup B) = P(\Omega) = 1$

$AB = \{b\}$ do đó $P(AB) = P(b) = 0,3$.

Ví dụ 1.25: Xét mạng gồm 4 chuyển mạch như sơ đồ sau. Mỗi vị trí chuyển mạch đều có hai trạng thái đóng hoặc mở đồng khả năng. Tính xác suất đoạn mạch giữa M và N ở trạng thái đóng.



Hình 1.6

Giải: Đặt A_k là biến cố "chuyển mạch s_k ở trạng thái đóng". Gọi A là biến cố "đoạn mạch giữa M và N ở trạng thái đóng".

$$A = A_1 \cup \left[A_2 \left(A_3 \cup A_4 \right) \right] = A_1 \cup \left(A_2 A_3 \right) \cup \left(A_2 A_4 \right).$$

Áp dụng công thức (1.11b) ta có:

$$\begin{aligned} P(A) &= P\left[A_1 \cup \left(A_2 A_3 \right) \cup \left(A_2 A_4 \right) \right] \\ &= P\left(A_1 \right) + P\left(A_2 A_3 \right) + P\left(A_2 A_4 \right) - P\left[A_1 \left(A_2 A_3 \right) \right] \\ &\quad - P\left[A_1 \left(A_2 A_4 \right) \right] - P\left[\left(A_2 A_3 \right) \left(A_2 A_4 \right) \right] + P\left(A_1 A_2 A_3 A_4 \right). \end{aligned}$$

Mỗi chuyển mạch s_k có 2 trạng thái, vậy đoạn mạch giữa *M* và *N* có 16 trạng thái đồng khả năng. Nếu chuyển mạch ở trạng thái đóng ta ký hiệu 1 và ở trạng thái mở ta ký hiệu 0. Ta có thể liệt kê tất cả các trường hợp có thể và sự xuất hiện các biến cố theo bảng sau:

s_1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
s_2	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
s_3	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
s_4	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0

Do đó:

$$P(A_1)=\frac{8}{16},\ P(A_2A_3)=P(A_2A_4)=\frac{4}{16},$$

$$P(A_1A_2A_3)=P(A_1A_2A_4)=P(A_2A_3A_4)=\frac{2}{16},$$

$$P(A_1A_2A_3A_4)=\frac{1}{16}.$$

Vậy: $P(A)=\frac{8}{16}+2.\frac{4}{16}-3.\frac{2}{16}+\frac{1}{16}=\frac{11}{16}\approx 0,688.$

1.2.6 Nguyên lý xác suất lớn, xác suất nhỏ

Biến cố không thể có xác suất bằng 0. Tuy nhiên một biến cố có xác suất gần bằng 0 vẫn có thể xảy ra khi thực hiện một số lớn các phép thử. Qua thực nghiệm và quan sát thực tế, người ta thấy rằng các biến cố có xác suất nhỏ sẽ không xảy ra khi ta chỉ thực hiện một phép thử hay một vài phép thử. Từ đó ta thừa nhận nguyên lý sau đây, gọi là "Nguyên lý xác suất nhỏ": *Nếu một biến cố có xác suất rất nhỏ thì thực tế có thể cho rằng trong một phép thử biến cố đó sẽ không xảy ra.*

Chẳng hạn mỗi chiếc máy bay đều có một xác suất rất nhỏ bị xảy ra tai nạn. Nhưng trên thực tế ta vẫn không từ chối đi máy bay vì tin tưởng rằng trong chuyến bay ta đi sự kiện máy bay rơi không xảy ra.

Hiển nhiên việc quy định một mức xác suất thế nào được gọi là nhỏ sẽ phụ thuộc vào từng bài toán cụ thể. Chẳng hạn nếu xác suất để máy bay rơi là 0,01 thì xác suất đó chưa thể được coi là nhỏ. Song nếu xác suất một chuyến tàu khởi hành chậm là 0,01 thì có thể coi rằng xác suất này là nhỏ.

Mức xác suất nhỏ này được gọi là *mức ý nghĩa*. Nếu α là mức ý nghĩa thì số $\beta = 1 - \alpha$ gọi là *độ tin cậy*. Khi dựa trên nguyên lý xác suất nhỏ ta khẳng định rằng: "Biến cố A có xác suất nhỏ (tức là $P(A) \le \alpha$) sẽ không xảy ra trên thực tế" thì độ tin cậy của kết luận trên là β. Tính đúng đắn của kết luận chỉ xảy ra trong $100 \cdot \beta\%$ trường hợp.

Tương tự như vậy ta có thể đưa ra "Nguyên lý xác suất lớn": "*Nếu biến cố A có xác suất gần bằng 1 thì trên thực tế có thể cho rằng biến cố đó sẽ xảy ra trong một phép thử*". Cũng như trên, việc quy định một mức xác suất thế nào được gọi là lớn sẽ tùy thuộc vào từng bài toán cụ thể.

1.3 XÁC SUẤT CÓ ĐIỀU KIỆN

1.3.1 Định nghĩa và các tính chất của xác suất có điều kiện

Định nghĩa 1.4: *Xác suất của biến cố B được tính trong điều kiện biến cố A xảy ra được gọi là xác suất của B với điều kiện A. Ký hiệu $P(B|A)$.*

Tính chất

- Nếu $P(A) > 0$ thì: $P(B|A) = \dfrac{P(AB)}{P(A)}$ (1.13)

- Khi cố định A với $P(A) > 0$ thì xác suất có điều kiện $P(B|A)$ có tất cả các tính chất của xác suất thông thường (công thức (1.7); (1.12)) đối với biến cố B.

Chẳng hạn:

$$\begin{aligned} &P(\overline{B}|A) = 1 - P(B|A), \\ &P(B_1 \cup B_2 | A) = P(B_1|A) + P(B_2|A) - P(B_1 B_2 | A) \ldots \end{aligned} \quad (1.14)$$

Ví dụ 1.26: Gieo đồng thời hai con xúc sắc (6 mặt) cân đối. Tính xác suất để tổng số chấm xuất hiện trên hai con xúc sắc ≥ 10 biết rằng ít nhất một con đã ra chấm 5.

Giải: Gọi A là biến cố "ít nhất một con ra chấm 5".

$$P(A) = 1 - P\left(\overline{A}\right) = 1 - \left(\frac{5}{6}\right)^2 = \frac{11}{36}.$$

Gọi B là biến cố "tổng số chấm trên hai con ≥ 10"

Biến cố AB có 3 kết quả thuận lợi là (5,6) , (5,5), (6,5).

Vậy $P(AB) = \frac{3}{36} \Rightarrow P\left(B|A\right) = \frac{3}{36} \Big/ \frac{11}{36} = \frac{3}{11}$.

Ví dụ 1.27: Xét phép thử gieo đồng xu liên tiếp 3 lần ở ví dụ 1.23

Gọi A là biến cố lần thứ nhất ra mặt sấp.

B là biến cố lần thứ hai ra mặt ngửa.

C là biến cố số lần mặt sấp xuất hiện nhiều hơn hoặc bằng số lần mặt ngửa. Tính *P(B/A), P(C/A)*.

Giải: Từ sơ đồ cây ở hình 1.5 ta suy ra:

$$P(A) = \frac{1}{2};\ P(AB) = \frac{1}{4} \Rightarrow P(B \mid A) = \frac{1}{4} \Big/ \frac{1}{2} = \frac{1}{2}$$

$$AC = \{\omega_1, \omega_2, \omega_3\} \Rightarrow P(AC) = \frac{3}{8} \Rightarrow P(C \mid A) = \frac{3}{8} \Big/ \frac{1}{2} = \frac{3}{4}$$

Ví dụ 1.28: Xét phép thử gieo liên tiếp 2 lần con xúc sắc 4 mặt trong ví dụ 1.10. Gọi X , Y lần lượt là số chấm xuất hiện khi gieo lần thứ nhất và lần thứ hai. Ta tính xác suất có điều kiện $P(B \mid A)$ trong đó

$$A = \{\max(X,Y) = m\},\quad B = \{\min(X,Y) = 2\}$$

và m nhận một trong các giá trị 1, 2, 3, 4.

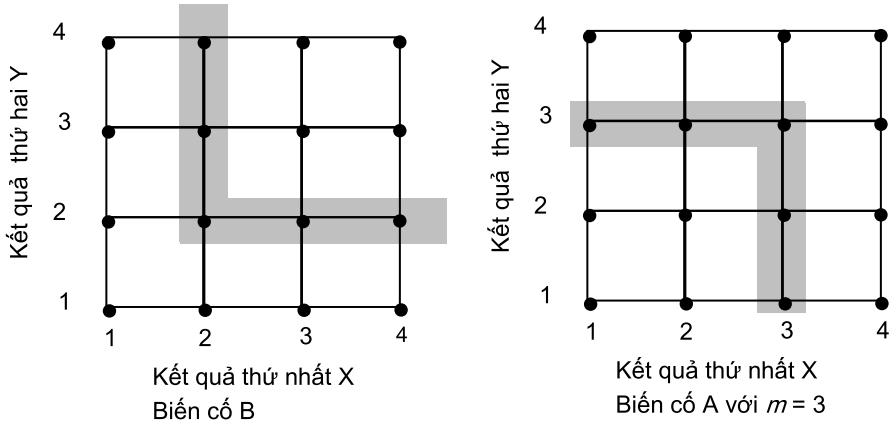
Giải: Có thể biểu diễn không gian mẫu của phép thử và các biến cố tương ứng trong hình 1.7.

Biến cố B được đánh dấu ở đồ thị trái của hình 1.7; Biến cố A được đánh dấu ở đồ thị phải của hình 1.7; Biến cố AB là phần giao của hai đồ thị trên của A và B.

Từ hình 1.7 ta được: $P(B) = \frac{5}{16}$;

$$P(AB)=\begin{cases}2/16 & \text{nếu } m=3 \text{ hoặc } m=4\\ 1/16 & \text{nếu } m=2\\ 0 & \text{nếu } m=1\end{cases}$$

$$P(A\,|\,B)=\begin{cases}2/5 & \text{nếu } m=3 \text{ hoặc } m=4\\ 1/5 & \text{nếu } m=2\\ 0 & \text{nếu } m=1\end{cases}$$



Hình 1.7: Phép thử gieo liên tiếp 2 lần xúc sắc 4 mặt

Ví dụ 1.29: Có hai phân xưởng của nhà máy sản xuất cùng một loại sản phẩm. Phân xưởng I sản xuất được 1000 sản phẩm trong đó có 100 phế phẩm. Phân xưởng II sản xuất được 2000 sản phẩm trong đó có 150 phế phẩm. Lấy ngẫu nhiên một sản phẩm để kiểm tra và đó là phế phẩm. Tính xác suất phế phẩm này do phân xưởng thứ I sản xuất.

Giải: Gọi B là biến cố sản phẩm được chọn để kiểm tra là phế phẩm. Gọi A là biến cố sản phẩm được chọn để kiểm tra do phân xưởng I sản xuất. Ta cần tính xác suất có điều kiện $P(A\,|\,B)$.

Biến cố AB có 100 kết quả thuận lợi đồng khả năng, do đó:

$$P(AB)=\frac{100}{3000}=\frac{1}{30}\ .$$

Trong 3000 sản phẩm sản xuất ra có 250 phế phẩm, do đó:

$$P(B)=\frac{250}{3000}=\frac{1}{12}.$$

Áp dụng công thức (1.13) ta được:

$$P(A \mid B) = \frac{1/30}{1/12} = 2/5 = 0,4 .$$

1.3.2 Quy tắc nhân xác suất

1.3.2.1 Trường hợp độc lập

- Nếu A, B là hai biến cố độc lập thì xác suất của biến cố B không phụ thuộc vào A có xảy ra hay không (xem mục 1.1.3–g), nghĩa là $P(B \mid A) = P(B)$. Theo công thức (1.13) ta có:

$$P(AB) = P(A)P(B) \tag{1.15}$$

- Nếu $\{A_1, A_2, ..., A_n\}$ là các biến cố độc lập thì:

$$P(A_1 A_2 ... A_n) = P(A_1)P(A_2)...P(A_n) \tag{1.16}$$

Thông thường tính độc lập của các biến cố được suy ra từ ý nghĩa thực tế. Chẳng hạn nếu A và B là biến cố xạ thủ 1, 2 bắn trúng mục tiêu thì A, B là hai biến cố độc lập (xem ví dụ 1.8).

1.3.2.2 Trường hợp tổng quát

- Với hai biến cố A, B bất kỳ, áp dụng công thức (1.13) ta có

$$P(AB) = P(A)P(B \mid A) \tag{1.17}$$

- Với n biến cố bất kỳ $A_1, A_2, ..., A_n$:

$$\begin{aligned} &P(A_1 A_2 ... A_n) \\ &= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 A_2)...P(A_n \mid A_1 A_2 ... A_{n-1}) \end{aligned} \tag{1.18}$$

Ví dụ 1.30: Túi I chứa 3 bi trắng, 7 bi đỏ, 15 bi xanh.

Túi II chứa 10 bi trắng, 6 bi đỏ, 9 bi xanh.

Từ mỗi túi lấy ngẫu nhiên 1 bi. Tìm xác suất để 2 bi được rút từ 2 túi là cùng màu.

Giải:

Gọi A_t, $A_đ$, A_x lần lượt là biến cố bi được rút từ túi I là trắng, đỏ, xanh.

B_t, $B_đ$, B_x lần lượt là biến cố bi được rút từ túi II là trắng, đỏ, xanh.

Các biến cố A_t, $A_đ$, A_x xung khắc, B_t, $B_đ$, B_x xung khắc;

Các biến cố A_t, $A_đ$, A_x độc lập với các biến cố B_t, $B_đ$, B_x.

Biến cố 2 bi được rút cùng màu là $A_tB_t \cup A_đB_đ \cup A_xB_x$

Vậy xác suất cần tìm:

$$P\left(A_tB_t \cup A_đB_đ \cup A_xB_x\right) = P\left(A_tB_t\right) + P\left(A_đB_đ\right) + P\left(A_xB_x\right)$$

$$= P\left(A_t\right)P\left(B_t\right) + P\left(A_đ\right)P\left(B_đ\right) + P\left(A_x\right)P\left(B_x\right)$$

$$= \frac{3}{25}\frac{10}{25} + \frac{7}{25}\frac{6}{25} + \frac{15}{25}\frac{9}{25} = \frac{207}{625} \approx 0{,}331$$

Ví dụ 1.31: Hai máy bay ném bom 1 mục tiêu, mỗi máy bay ném 1 quả với xác suất trúng mục tiêu tương ứng là $0,7$ và $0,8$. Tìm xác suất để mục tiêu bị trúng bom.

Giải: Gọi A_1, A_2 lần lượt tương ứng là biến cố "máy bay thứ nhất và máy bay thứ hai ném trúng mục tiêu". A là biến cố "mục tiêu bị đánh trúng".

Rõ ràng $A = A_1 \cup A_2$ và A_1, A_2 độc lập.

Do đó:

$$P(A) = P\left(A_1 \cup A_2\right) = P\left(A_1\right) + P\left(A_2\right) - P\left(A_1\right)P\left(A_2\right)$$
$$= 0,7 + 0,8 - 0,7.0,8 = 0,96$$

Ví dụ 1.32: Rút ngẫu nhiên 2 quân bài từ cỗ bài tú lơ khơ. Tính xác suất cả 2 quân bài rút được là 2 con át.

Giải: : Gọi A_1, A_2 lần lượt tương ứng là biến cố lần thứ nhất và lần thứ hai rút được con át.

$$P(A_1A_2) = P(A_1)P(A_2 \mid A_1) = \frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221}.$$

Ví dụ 1.33: Một hộp đựng 100 con chíp bán dẫn trong đó có 20 chíp là phế phẩm. Lấy lần lượt ngẫu nhiên không hoàn lại 2 chíp bán dẫn ở trong hộp.

a) Tính xác suất con chíp lấy được lần đầu là phế phẩm.

b) Tính xác suất con chíp lấy được lần thứ hai là phế phẩm biết rằng con chíp lấy lần đầu cũng là phế phẩm.

c) Tính xác suất cả hai con chíp lấy được đều là phế phẩm.

***Giải*:** a) Gọi A_1 là biến cố con chíp lấy được lần đầu là phế phẩm, ta có:

$$P(A_1) = \frac{20}{100} = 0,2$$

b) Gọi A_2 là biến cố con chíp lấy được lần thứ hai là phế phẩm. Vậy xác suất con chíp lấy được lần thứ hai là phế phẩm biết rằng con chíp lấy lần đầu cũng là phế phẩm:

$$P(A_2 \mid A_1) = \frac{19}{99} = 0,192$$

c) $P(A_1A_2) = P(A_1)P(A_2 \mid A_1) = \frac{20}{100} \cdot \frac{19}{99} = 0,0384$

Ví dụ 1.34: Một thủ kho có một chùm chìa khóa gồm 9 chiếc, bề ngoài chúng giống hệt nhau nhưng trong đó chỉ có đúng 2 chiếc mở được kho. Anh ta thử ngẫu nhiên từng chìa (chìa nào không trúng thì bỏ ra). Tính xác suất để mở được kho ở lần thứ ba.

Giải: Ký hiệu A_i là biến cố "thử đúng chìa ở lần thứ i", $i = 1,...,8$

Ký hiệu B là biến cố "mở được kho ở lần thử thứ ba".

Ta có: $B = \overline{A_1}\,\overline{A_2}\,A_3$.

Vậy xác suất cần tìm là:

$$P\left(\overline{A_1}\,\overline{A_2}\,A_3\right) = P\left(\overline{A_1}\right)P\left(\overline{A_2}\,\middle|\,\overline{A_1}\right)P\left(A_3\middle|\,\overline{A_1}\,\overline{A_2}\right)$$

Có thể tính được: $P\left(\overline{A_1}\right)=\frac{7}{9}$, $P\left(\overline{A_2}\,\middle|\,\overline{A_1}\right)=\frac{6}{8}$, $P\left(A_3\middle|\overline{A_1}\,\overline{A_2}\right)=\frac{2}{7}$

Do đó: $P\left(\overline{A_1}\,\overline{A_2}\,A_3\right)=\frac{7}{9}\frac{6}{8}\frac{2}{7}=\frac{1}{6}$

Ví dụ 1.35: Rút lần lượt ngẫu nhiên không hoàn lại 3 quân bài từ cỗ bài tú lơ khơ. Tính xác suất trong các trường hợp sau:

a) Cả 3 quân bài rút được không phải là quân bích.

b) Lần thứ nhất rút được không phải quân bích và lần thứ hai rút được quân bích.

c) Hai lần đầu rút được không phải quân bích và lần thứ ba rút được quân bích.

Giải: : Gọi A_1, A_2, A_3 lần lượt tương ứng là biến cố lần thứ nhất, lần thứ hai và lần thứ ba rút được quân bài không phải là bích.

a) Biến cố cả 3 quân bài rút được không phải là quân bích là $A_1A_2A_3$.

Vậy xác suất cần tìm là:

$$P(A_1A_2A_3)=P(A_1)P(A_2\mid A_1)P(A_3\mid A_1A_2).$$

$$P(A_1)=\frac{39}{52},\ P(A_2\mid A_1)=\frac{38}{51},\ P(A_3\mid A_1A_2)=\frac{37}{50}.$$

$$P(A_1A_2A_3)=\frac{39}{52}\cdot\frac{38}{51}\cdot\frac{37}{50}.$$

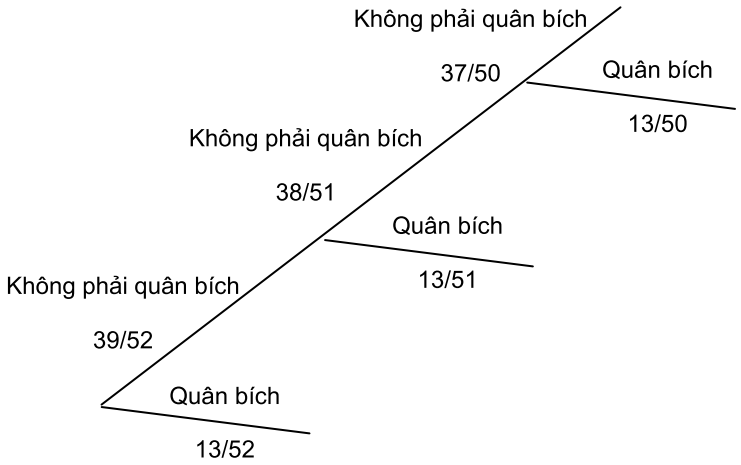
b) Xác suất lần thứ nhất rút được không phải quân bích và lần thứ hai rút được quân bích là:

$$P\left(A_1\overline{A_2}\right)=P\left(A_1\right)P\left(\overline{A_2}\mid A_1\right)=\frac{39}{52}\cdot\frac{13}{51}$$

c) Xác suất hai lần đầu rút được không phải quân bích và lần thứ ba rút được quân bích là

$$P\left(A_1A_2\overline{A_3}\right)=P\left(A_1\right)P\left(A_2\mid A_1\right)P\left(\overline{A_3}\mid A_1A_2\right)=\frac{39}{52}\cdot\frac{38}{51}\cdot\frac{13}{50}$$

Tương tự ví dụ 1.23 ta có thể biểu diễn các biến cố và xác suất tương ứng của phép thử rút liên tiếp 3 quân bài dưới dạng sơ đồ cây.



Hình 1.7: Phép thử gieo liên tiếp 2 lần xúc sắc 4 mặt

1.3.3 Công thức xác suất đầy đủ

Định lý 1.2: Giả sử $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ các biến cố. Khi đó, với mọi biến cố B của cùng một phép thử ta có:

$$P(B) = \sum_{i=1}^{n} P(A_i B) = \sum_{i=1}^{n} P(A_i) P\left(B \mid A_i\right) \qquad (1.19)$$

Ví dụ 1.36: Một túi đựng 4 bi trắng và 6 bi đen. Người thứ nhất lấy ngẫu nhiên từ túi 3 bi (không hoàn lại), người thứ hai lấy tiếp 2 bi. Tính xác suất để người thứ hai lấy được 1 bi trắng.

Giải: Gọi lần lượt A_0, A_1, A_2, A_3 là biến cố người thứ nhất lấy được 0, 1, 2, 3 bi trắng.

Gọi B là biến cố người thứ hai lấy được 1 bi trắng.

Ta có: $P(A_0) = \dfrac{C_6^3}{C_{10}^3} = \dfrac{1}{6}$, $P(A_1) = \dfrac{C_4^1 C_6^2}{C_{10}^3} = \dfrac{1}{2}$,

$P(A_2) = \dfrac{C_4^2 C_6^1}{C_{10}^3} = \dfrac{3}{10}$, $P(A_3) = \dfrac{C_4^3}{C_{10}^3} = \dfrac{1}{30}$.

$$P(B|A_0)=\frac{C_4^1C_3^1}{C_7^2}=\frac{12}{21},\ P(B|A_1)=\frac{C_3^1C_4^1}{C_7^2}=\frac{12}{21},$$

$$P(B|A_2)=\frac{C_2^1C_5^1}{C_7^2}=\frac{10}{21},\ \ P(B|A_3)=\frac{C_1^1C_6^1}{C_7^2}=\frac{6}{21}.$$

Do đó $P(B)=\frac{1}{6}\cdot\frac{12}{21}+\frac{1}{2}\cdot\frac{12}{21}+\frac{3}{10}\cdot\frac{10}{21}+\frac{1}{30}\cdot\frac{6}{21}=\frac{56}{105}.$

Ví dụ 1.37: Gieo xúc sắc 4 mặt (xem ví dụ 1.10). Nếu mặt 1 chấm hoặc 2 chấm xuất hiện ta gieo tiếp lần nữa và ngừng nếu ngược lại. Tính xác suất để tổng số chấm xuất hiện ít nhất là 4.

Giải: Gọi A_k là biến cố lần gieo thứ nhất xuất hiện k chấm, ta có:

$$P(A_k)=\frac{1}{4} \text{ với mọi } k=1,2,3,4.$$

Gọi B là biến cố tổng số chấm xuất hiện ít nhất là 4.

Giả sử biến cố A_1 xảy ra, khi đó tổng số chấm ít nhất là 4 khi kết quả của lần gieo thứ hai là 3 hoặc 4. Tương tự, nếu biến cố A_2 xảy ra, khi đó tổng số chấm ít nhất là 4 khi kết quả của lần gieo thứ hai là 2, 3 hoặc 4. Vậy:

$$P(B\,|\,A_1)=\frac{1}{2},\qquad P(B\,|\,A_2)=\frac{3}{4}$$

Nếu biến cố A_3 hoặc A_4 xảy ra thì dừng lại không gieo tiếp lần thứ hai, do đó:

$$P(B\,|\,A_3)=0,\qquad P(B\,|\,A_4)=1.$$

Áp dụng công thức xác suất đầy đủ ta được:

$$P(B)=\frac{1}{4}\cdot\frac{1}{2}+\frac{1}{4}\cdot\frac{3}{4}+\frac{1}{4}\cdot 0+\frac{1}{4}\cdot 1=\frac{9}{16}.$$

1.3.4 Công thức Bayes

Định lý 1.3: Giả sử $\{A_1, A_2, ..., A_n\}$ là một hệ đầy đủ các biến cố. Khi đó, với mọi biến cố B của cùng một phép thử và $P(B)>0$ ta có:

$$P\left(A_k \middle| B\right) = \frac{P(A_k B)}{P(B)} = \frac{P(A_k)P\left(B \middle| A_k\right)}{\sum_{i=1}^{n} P(A_i)P\left(B \middle| A_i\right)} \tag{1.20}$$

Nhận xét 1.2:

Trong thực tế các xác suất $\{P(A_1), P(A_2), ..., P(A_n)\}$ đã biết và được gọi là các *xác suất tiền nghiệm*. Sau khi quan sát biết được biến cố B xảy ra, các xác suất của A_k được tính trên thông tin này (xác suất có điều kiện $P(A_k|B)$) được gọi là *xác suất hậu nghiệm*. Vì vậy công thức Bayes còn được gọi là công thức xác suất hậu nghiệm.

Ví dụ 1.38: Xét kênh viễn thông nhị phân được biểu diễn như sơ đồ hình 1.9.

Đầu vào của kênh ký hiệu là X và giả thiết rằng chỉ có hai trạng thái 0 và 1, tương tự đầu ra ký hiệu là Y và cũng chỉ có hai trạng thái 0 và 1. Do bị nhiễu kênh nên đầu vào 0 có thể chuyển thành đầu ra là 1 và ngược lại.

Gọi là X_0 biến cố "X có trạng thái 0" và X_1 là biến cố "X có trạng thái 1".

Gọi là Y_0 biến cố "đầu ra Y có trạng thái 0" và Y_1 là biến cố "đầu ra Y có trạng thái 1". Khi đó $\{X_0, X_1\}$ và $\{Y_0, Y_1\}$ là hai hệ đầy đủ.

Kênh được đặc trưng bởi các xác suất chuyển p_0, q_0, p_1 và q_1, trong đó:

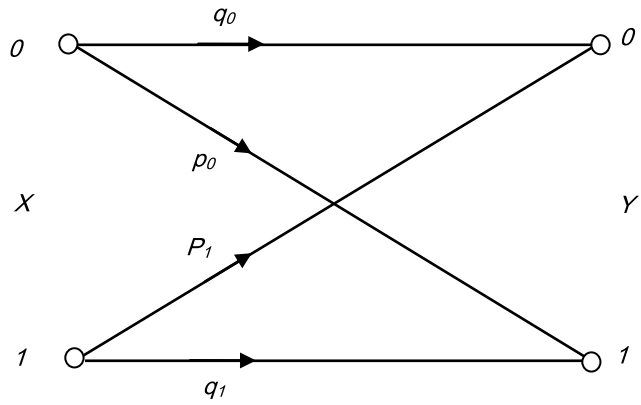
$$p_0 = P\left(Y_1 \middle| X_0\right) \text{ và } p_1 = P\left(Y_0 \middle| X_1\right)$$

$$q_0 = P\left(Y_0 \middle| X_0\right) \text{ và } q_1 = P\left(Y_1 \middle| X_1\right)$$

Ta thấy rằng:

$$p_0 + q_0 = 1 = p_1 + q_1.$$

Giả sử $P\left(X_0\right) = 0,5$, $p_0 = 0,1$ và $p_1 = 0,2$.



Hình 1.9

a) Tìm xác suất đầu ra của kênh là 0 và xác suất đầu ra của kênh là 1.

b) Giả sử đầu ra của kênh nhận được là 0. Tìm xác suất nhận đúng tín hiệu đầu vào.

c) Tính xác suất lỗi P_e

Giải:

$$P(X_1) = 1 - P(X_0) = 0,5;$$

$$q_0 = 1 - p_0 = 1 - 0,1 = 0,9;$$

$$q_1 = 1 - p_1 = 1 - 0,2 = 0,8$$

a) Áp dụng công thức xác suất đầy đủ với hệ đầy đủ $\{X_0, X_1\}$ ta được:

$$P(Y_0) = P(X_0)P(Y_0|X_0) + P(X_1)P(Y_0|X_1) = 0,5 \times 0,9 + 0,5 \times 0,2 = 0,55$$

$$P(Y_1) = P(X_0)P(Y_1|X_0) + P(X_1)P(Y_1|X_1) = 0,5 \times 0,1 + 0,5 \times 0,8 = 0,45$$

b) Áp dụng công thức Bayes ta có:

$$P(X_0|Y_0) = \frac{P(X_0)P(Y_0|X_0)}{P(Y_0)} = \frac{0,5 \times 0,9}{0,55} = 0,818$$

c) Xác suất lỗi

$$P_e = P(X_0Y_1 \cup X_1Y_0) = P(X_0)P(Y_1|X_0) + P(X_1)P(Y_0|X_1)$$
$$= 0,5 \times 0,1 + 0,5 \times 0,2 = 0,15$$

Ví dụ 1.39: Một nhà máy có ba phân xưởng I, II, III cùng sản xuất ra một loại sản phẩm. Phân xưởng I, II, III sản xuất tương ứng 36%, 34%, 30% sản lượng của nhà máy, với tỷ lệ phế phẩm tương ứng là 0,12; 0,1; 0,08.

a) Tìm tỷ lệ phế phẩm chung của nhà máy.

b) Lấy ngẫu nhiên một sản phẩm kiểm tra và đó là phế phẩm. Tính xác suất để phế phẩm đó là do phân xưởng I sản xuất.

Giải: Lấy ngẫu nhiên một sản phẩm của nhà máy để kiểm tra. Gọi B là biến cố "sản phẩm kiểm tra là phế phẩm".

Gọi A_1, A_2, A_3 lần lượt là biến cố sản phẩm lấy ra kiểm tra do phân xưởng I, II, III sản xuất.

Theo giả thiết ta có: hệ 3 biến cố $\{A_1, A_2, A_3\}$ đầy đủ (xem ví dụ 1.7).

$$P(A_1) = 0,36;\ P(A_2) = 0,34;\ P(A_3) = 0,30.$$

$$P(B|A_1) = 0,12;\ P(B|A_2) = 0,10;\ P(B|A_3) = 0,08.$$

a. Xác suất của biến cố B cũng là tỉ lệ phế phẩm chung của nhà máy. Áp dụng công thức xác suất đầy đủ (1.19) ta có:

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) = 0,1012$$

b. Áp dụng công thức Bayes ta được:

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(B)} = \frac{0,36 \cdot 0,12}{0,1012} = 0,427$$

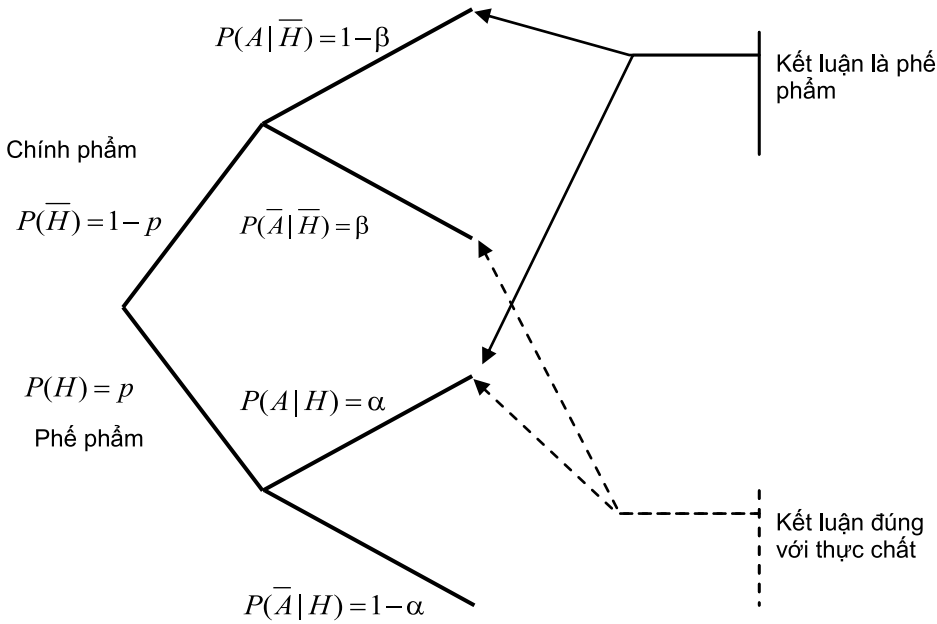
Ví dụ 1.40: Người ta dùng một thiết bị để kiểm tra một loại sản phẩm nhằm xác định sản phẩm có đạt yêu cầu không. Biết rằng sản phẩm có tỉ lệ phế phẩm là $p\%$. Thiết bị có khả năng phát hiện đúng

sản phẩm là phế phẩm với xác suất α và phát hiện đúng sản phẩm đạt chất lượng với xác suất β. Kiểm tra ngẫu nhiên một sản phẩm, tìm xác suất sao cho sản phẩm này:

a) Được kết luận là phế phẩm (biến cố A).

b) Được kết luận là đạt chất lượng thì lại là phế phẩm.

c) Được kết luận đúng với thực chất của nó.



Hình 1.10: Sơ đồ cây xác suất đầy đủ

Giải: Gọi H là biến cố "sản phẩm được chọn là phế phẩm". Theo giả thiết ta có:

$$P(H)=p,\ P\left(A\middle|H\right)=\alpha,\ P\left(\overline{A}\middle|\overline{H}\right)=\beta.$$

a) Áp dụng công thức đầy đủ cho hệ đầy đủ $\left\{H,\overline{H}\right\}$ ta có:

$$P(A)=P(H)P\left(A\middle|H\right)+P\left(\overline{H}\right)P\left(A\middle|\overline{H}\right)=p\alpha+(1-p)(1-\beta)$$

b) $P(\overline{A}) = P(H)P\left(\overline{A}\,\middle|\,H\right) + P\left(\overline{H}\right)P\left(\overline{A}\,\middle|\,\overline{H}\right) = p(1-\alpha) + (1-p)\beta$

$$\Rightarrow P\left(H\,\middle|\,\overline{A}\right) = \frac{P\left(H\overline{A}\right)}{P\left(\overline{A}\right)} = \frac{P(H)P\left(\overline{A}\,\middle|\,H\right)}{P\left(\overline{A}\right)} = \frac{p(1-\alpha)}{p(1-\alpha)+(1-p)\beta}$$

c) $P\left(AH\right) + P\left(\overline{A}\,\overline{H}\right) = P(H)P\left(A\,\middle|\,H\right) + P\left(\overline{H}\right)P\left(\overline{A}\,\middle|\,\overline{H}\right) = p\alpha + (1-p)\beta$

Ta có thể biểu diễn kết quả dưới dạng cây biểu đồ như hình 1.10.

1.4 DÃY PHÉP THỬ BERNOULLI

Định nghĩa 1.5: *Dãy các phép thử lặp lại được, độc lập, trong mỗi phép thử chỉ xét 2 kết cục:* A, $\overline{A}$ *và xác suất xuất hiện của biến cố* A *không đổi* $P(A) = p$, $(0 < p < 1)$ *được gọi là dãy phép thử Bernoulli.*

p gọi là xác suất thành công trong mỗi lần thử.

Kí hiệu H_k là biến cố "A xuất hiện ra đúng k lần trong n phép thử".

Đặt $P_n(k;p) = P(H_k)$.

Định lý 1.4: Xác suất của biến cố "A xuất hiện ra đúng k lần trong n phép thử":

$$P_n(k;p) = C_n^k p^k (1-p)^{n-k};\ k = 0,1,...,n \qquad (1.21)$$

Chứng minh: H_k là tổng của C_n^k các biến cố xung khắc từng đôi nhận được bằng cách hoán vị các chữ A và $\overline{A}$ trong biến cố tích sau:

$$\underbrace{A...A}_{k \text{ lần}}\ \underbrace{\overline{A}\ ...\ \overline{A}}_{n-k \text{ lần}}$$

Mỗi biến cố này có xác suất $P(\underbrace{A...A}_{k \text{ lần}}\underbrace{\overline{A}\ ...\ \overline{A}}_{n-k \text{ lần}}) = p^k(1-p)^{n-k}$

Vậy: $P_n(k;p) = C_n^k p^k (1-p)^{n-k}$.

Ta cần tìm giá trị $k, 0 \le k \le n$ sao cho xác suất $P_n(k;p)$ đạt giá trị lớn nhất.

Định lý 1.5:

(i) $P_n(k;p)=\dfrac{(n-k+1)p}{kq}P_n(k-1;p)$ (1.22)

(ii) Khi k tăng từ 0 đến n thì $P_n(k;p)$ mới đầu tăng sau đó giảm và đạt giá trị lớn nhất tại $k=m$ thoả mãn:

$$P_{\max}=P_n(m;p);\ (n+1)p-1\le m\le (n+1)p \qquad (1.23a)$$

Như vậy:

- Khi $(n+1)p$ không nguyên thì $m=\left[(n+1)p\right]$ (là phần nguyên của $(n+1)p$).

- Khi $(n+1)p$ nguyên thì $m=(n+1)p-1$ hoặc $m=(n+1)p$

$$P_{\max}=P_n(m-1;p)=P_n(m;p) \qquad (1.23b)$$

Chứng minh:

$\dfrac{P_n(k;p)}{P_n(k-1;p)}=\dfrac{\dfrac{n!}{k!(n-k)!}p^kq^{n-k}}{\dfrac{n!}{(k-1)!(n-k+1)!}p^{k-1}q^{n-k+1}}=\dfrac{(n-k+1)p}{kq}$, từ đó có (1.22).

(1.22) $\Rightarrow$ $\dfrac{P_n(k;p)}{P_n(k+1;p)}=\dfrac{(k+1)(1-p)}{(n-k)p}$

Do đó: $\dfrac{P_n(k;p)}{P_n(k+1;p)}<1\Leftrightarrow k+1<(n+1)p$

Vậy: $P_n(k;p)<P_n(k+1;p)$ khi $k<(n+1)p-1$

$\Rightarrow P_n(k;p)<P_n(m;p),\ \forall k<(n+1)p-1.$

và $P_n(k;p)>P_n(k+1;p)$ khi $k\ge(n+1)p$

$\Rightarrow P_n(k;p)<P_n(m;p),\ \forall k>(n+1)p$, trong đó m là số tự nhiên thỏa mãn $(n+1)p-1\le m\le(n+1)p$.

Khi $m=(n+1)p$ thì:

$$\frac{P_n(m-1;p)}{P_n(m;p)}=\frac{(n+1)(1-p)p}{\big(n-(n+1)p+1\big)p}=\frac{(n+1)(1-p)p}{\big(n+1-(n+1)p\big)p}=1$$

$$\Rightarrow P_n(m-1;p)=P_n(m;p).$$

Định nghĩa 1.6: *m xác định bởi công thức (1.23a) hoặc (1.23b) được gọi là số lần xuất hiện có khả năng nhất hay giá trị có khả năng xảy ra lớn nhất.*

Ví dụ 1.41: Bắn 7 viên đạn vào bia. Xác suất trúng đích của mỗi viên là $0,6$. Tìm xác suất trong các trường hợp sau:

a) Có đúng 3 viên trúng bia.

b) Có ít nhất 6 viên trúng bia.

c) Có ít nhất 1 viên trúng bia.

d) Tìm số viên đạn trúng bia có khả năng lớn nhất.

Giải: Có thể xem bắn mỗi viên đạn vào bia là thực hiện một phép thử Bernoulli mà xác suất thành công của phép thử là xác suất bắn trúng bia, theo giả thiết là 0,6. Bắn 7 viên là thực hiện 7 lần phép thử. Vậy:

a) Xác suất để có đúng 3 viên trúng bia là:

$$P_7(3;0,6)=C_7^3(0,6)^3(0,4)^4=0,1935$$

b) Xác suất để có ít nhất 6 viên trúng bia là:

$$P_7(6;0,6)+P_7(7;0,6)=C_7^6(0,6)^6(0,4)+C_7^7(0,6)^7=0,1586$$

c) Xác suất để có ít nhất 1 viên trúng bia là:

$$1-P_7(0;0,6)=1-C_7^0(0,6)^0(0,4)^7=1-(0,4)^7=0,998$$

d) $(n+1)p = (7+1)(0,6) = 4,8$. Vậy số viên đạn có khả năng trúng bia nhất là 4.

Ví dụ 1.42: Tín hiệu thông tin được phát đi 3 lần độc lập nhau. Xác suất thu được mỗi lần là 0,4.

a) Tìm xác suất để nguồn thu nhận được thông tin đúng 2 lần.

b) Tìm xác suất để nguồn thu nhận được thông tin đó.

c) Nếu muốn xác suất thu được tin $\geq 0,9$ thì phải phát đi ít nhất bao nhiêu lần.

Giải: Có thể xem mỗi lần phát tin là một phép thử Bernoulli mà sự thành công của phép thử là nguồn thu nhận được tin, theo giả thiết xác suất thành công của mỗi lần thử là 0,4. Vậy:

a) Xác suất để nguồn thu nhận được thông tin đúng 2 lần là:

$$P_3(2;0,4) = C_3^2 (0,4)^2 (0,6) = 0,288$$

b) Xác suất để nguồn thu nhận được thông tin là:

$$P = 1 - P_3(0;0,4) = 1 - (0,6)^3 = 0,784$$

c) Xác suất để nguồn thu nhận được thông tin khi phát n lần là:

$$P = 1 - (0,6)^n .$$

Vậy nếu muốn xác suất thu được tin $\geq 0,9$ thì phải phát đi ít nhất n lần sao cho:

$$1 - (0,6)^n \geq 0,9 \Leftrightarrow (0,6)^n \leq 0,1 \Leftrightarrow n \geq \frac{\lg(0,1)}{\lg(0,6)} = \frac{-1}{-1+0,778} = 4,504$$

Chọn $n = 5$.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

1.1 Ta có thể có hai không gian mẫu Ω các biến cố sơ cấp cho cùng một phép thử $\mathscr{C}$?

Đúng ☐ Sai ☐ .

1.2 Các biến cố A và $\overline{A} \cup \overline{B}$ là xung khắc.

Đúng ☐ Sai ☐ .

1.3 Hai biến cố A và B xung khắc thì $P(A \cup B) = P(A) + P(B)$.

Đúng ☐ Sai ☐ .

1.4 Hệ hai biến cố $\left\{A, \overline{A}\right\}$ là một hệ đầy đủ.

Đúng ☐ Sai ☐.

1.5 Hai biến cố xung khắc là hai biến cố độc lập.

Đúng ☐ Sai ☐.

1.6 Các biến cố đối của hai biến cố độc lập cũng là độc lập.

Đúng ☐ Sai ☐.

1.7 Xác suất của tổng hai biến cố độc lập bằng tổng xác suất của hai biến cố này.

Đúng ☐ Sai ☐.

1.8 Xác suất của tích 2 biến cố xung khắc bằng tích 2 xác suất.

Đúng ☐ Sai ☐.

1.9 Khi áp dụng công thức xác suất đầy đủ để tính xác suất biến cố B dựa vào hệ đầy đủ $\{A_1, ..., A_n\}$ thì các biến cố B và $A_1, ..., A_n$ phải trong cùng một phép thử.

Đúng ☐ Sai ☐.

1.10 Cho $\Omega = \{a,b,c,d\}$ trong đó các biến cố sơ cấp là đồng khả năng. Biến cố $A = \{a,b\}$ và $B = \{a,c\}$ là phụ thuộc vì chúng cùng xảy ra khi biến cố sơ cấp a xảy ra.

Đúng ☐ Sai ☐.

1.11 Trong một hòm đựng 10 chi tiết đạt tiêu chuẩn và 5 chi tiết là phế phẩm. Lấy đồng thời 3 chi tiết. Tính xác suất:

a) Cả 3 chi tiết lấy ra thuộc loại đạt tiêu chuẩn.

b) Trong số 3 chi tiết lấy ra có 2 chi tiết đạt tiêu chuẩn.

1.12 Thang máy của một tòa nhà 7 tầng xuất phát từ tầng một với 3 khách. Tìm xác suất để:

a) Tất cả cùng ra ở tầng bốn.

b) Tất cả cùng ra ở một tầng.

c) Mỗi người ra một tầng khác nhau.

1.13 Một người gọi điện thoại cho bạn nhưng lại quên mất 3 chữ số cuối và chỉ nhớ rằng chúng khác nhau. Tìm xác suất để người đó quay số một lần được đúng số điện thoại của bạn.

1.14 Ta kiểm tra theo thứ tự một lô hàng có 10 sản phẩm. Mỗi sản phẩm thuộc một trong hai loại: Tốt hoặc Xấu. Ký hiệu A_k ($k = 1,...,10$) là biến cố chỉ sản phẩm kiểm tra thứ k thuộc loại xấu. Biểu diễn các biến cố sau theo A_k:

a) Cả 10 sản phẩm đều xấu.

b) Có ít nhất một sản phẩm xấu.

c) Có 6 sản phẩm kiểm tra đầu là tốt, các sản phẩm còn lại là xấu.

d) Có 6 sản phẩm kiểm tra đầu là xấu.

1.15 Hai người cùng bắn vào một mục tiêu. Khả năng bắn trúng của từng người là 0,8 và 0,9. Tìm xác suất:

a) Chỉ có một người bắn trúng mục tiêu.

b) Có người bắn trúng mục tiêu.

c) Cả hai người bắn trượt.

1.16 Cơ cấu chất lượng sản phẩm của nhà máy như sau: 40% sản phẩm là loại I, 50% sản phẩm là loại II, còn lại là phế phẩm. Lấy ngẫu nhiên một sản phẩm của nhà máy. Tính xác suất sản phẩm lấy ra là phế phẩm.

1.17 Có 1000 vé số trong đó có 20 vé trúng thưởng. Một người mua 30 vé, tìm xác suất để người đó trúng 5 vé.

1.18 Để được nhập kho, sản phẩm của nhà máy phải qua 3 vòng kiểm tra chất lượng độc lập nhau. Xác suất phát hiện ra phế phẩm ở các vòng lần lượt theo thứ tự là 0,8; 0,9 và 0,99. Tính xác suất phế phẩm được nhập kho.

1.19 Giả sử xác suất xuất hiện biến cố A trong mỗi lần thử $\mathscr{C}$ là p, $0 < p < 1$. Thực hiện phép thử $\mathscr{C}$ lặp lại, độc lập cho đến khi biến cố A xuất hiện và đếm số lần phải thực hiện phép thử $\mathscr{C}$ để lần đầu tiên biến cố A xuất hiện.

a) Tìm không gian mẫu Ω của phép thử.

b) Tính xác suất biến cố A xuất hiện lần đầu tiên ở lần thử thứ k.

c) Nghiệm lại rằng $P(\Omega) = 1$.

d) Gọi B là biến cố lần đầu tiên xuất hiện biến cố A với số lần thử là chẵn. Tính $P(A)$.

1.20 Một thủ kho có một chùm chìa khóa gồm 9 chiếc trông giống hệt nhau trong đó chỉ có một chiếc mở được kho. Anh ta thử ngẫu nhiên từng chìa khóa một, chiếc nào được thử thì không thử lại. Tính xác suất anh ta mở được cửa ở lần thử thứ 4.

1.21 Một lô hàng có 9 sản phẩm. Mỗi lần kiểm tra chất lượng lấy ngẫu nhiên 3 sản phẩm. Sau khi kiểm tra xong trả lại vào lô hàng. Tính xác suất để sau 3 lần kiểm tra lô hàng, tất cả các sản phẩm đều được kiểm tra.

1.22 Hai biến cố A, B có xác suất $P(A)=0,3$, $P(A\cup B)=0,65$. Giả sử A, B độc lập nhưng không xung khắc. Tính $P(B)$.

1.23 Giả sử hai biến cố A, B có xác suất $P(A)=1/2$, $P(B)=1/3$ và $P(AB)=1/4$. Hãy tính:

a) $P(A\,|\,B)$ b) $P(B\,|\,A)$ c) $P(A\cup B)$ d) $P(A\overline{B})$

e) $P(\overline{A}B)$ f) $P(\overline{B}\,|\,\overline{A})$ g) $P(\overline{A}\,|\,\overline{B})$ h) $P(\overline{A}\,\overline{B})$

1.24 Chọn ngẫu nhiên lần lượt không hoàn lại 2 số từ các số $\{0,1,...,9\}$. Tính xác suất số thứ hai chọn được là số 4.

1.25 Một nhà máy ôtô có ba phân xưởng I, II, III cùng sản xuất ra một loại pít-tông. Phân xưởng I, II, III sản xuất tương ứng 36%, 34%, 30% sản lượng của nhà máy, với tỷ lệ phế phẩm tương ứng là 0,12; 0,1; 0,08.

a) Tìm tỷ lệ phế phẩm chung của nhà máy.

b) Lấy ngẫu nhiên một sản phẩm kiểm tra và được sản phẩm là phế phẩm. Tính xác suất để phế phẩm đó là do phân xưởng I, II, III sản xuất.

1.26 Có bốn nhóm xạ thủ tập bắn. Nhóm thứ nhất có 5 người, nhóm thứ hai có 7 người, nhóm thứ ba có 4 người và nhóm thứ tư có 2 người. Xác suất bắn trúng đích của mỗi người trong nhóm thứ nhất, nhóm thứ hai, nhóm thứ ba và nhóm thứ tư theo thứ tự là 0,8; 0,7; 0,6 và 0,5. Chọn ngẫu nhiên một xạ thủ và biết rằng xạ thủ này bắn trượt. Hãy xác định xem xạ thủ này có khả năng ở trong nhóm nào nhất.

1.27 Bắn hai lần độc lập với nhau mỗi lần một viên đạn vào cùng một bia. Xác suất trúng đích của viên đạn thứ nhất là $0,7$ và của viên đạn thứ hai là $0,4$. Tìm xác suất để chỉ có một viên đạn trúng bia (biến cố A). Sau khi bắn, quan trắc viên báo có một vết đạn ở bia. Tìm xác suất để vết đạn đó là vết đạn của viên đạn thứ nhất.

1.28 Một nhà máy sản xuất một chi tiết của điện thoại di động có tỷ lệ sản phẩm đạt tiêu chuẩn chất lượng là 85%. Trước khi xuất xưởng người ta dùng một thiết bị kiểm tra để kết luận sản phẩm có đạt yêu cầu chất lượng hay không. Thiết bị có khả năng phát hiện đúng sản phẩm đạt tiêu chuẩn với xác suất là 0,9 và phát hiện đúng sản phẩm không đạt tiêu chuẩn với xác suất là 0,95. Tìm xác suất để 1 sản phẩm được chọn ngẫu nhiên sau khi kiểm tra:

a) Được kết luận là đạt tiêu chuẩn.

b) Được kết luận là đạt tiêu chuẩn thì lại không đạt tiêu chuẩn.

c) Được kết luận đúng với thực chất của nó.

1.29 Giả sử phép thử $\mathscr{C}$ có thể thực hiện lặp lại, độc lập. Không gian mẫu của phép thử là $\{A, B, C\}$ với xác suất tương ứng $P(A) = p$, $P(B) = q$, $P(C) = r$. Tính xác suất biến cố A xảy ra trước biến cố B.

1.30 Chứng minh rằng nếu $P(A \mid B) > P(A)$ thì $P(B \mid A) > P(B)$.

Chương 2

BIẾN NGẪU NHIÊN VÀ CÁC ĐẶC TRƯNG CỦA CHÚNG

GIỚI THIỆU

Trong chương này ta khảo sát các biến cố gắn với các giá trị nào đó, khi các giá trị này thay đổi ta được các biến ngẫu nhiên.

Khái niệm biến ngẫu nhiên (còn được gọi là đại lượng ngẫu nhiên) và các đặc trưng của chúng là những khái niệm rất quan trọng của lý thuyết xác suất.

Đối với biến ngẫu nhiên ta chỉ quan tâm đến vấn đề biến ngẫu nhiên này nhận một giá trị nào đó hoặc nhận giá trị trong một khoảng nào đó với xác suất bao nhiêu. Nói cách khác biến ngẫu nhiên X có thể được khảo sát thông qua hàm phân bố xác suất của nó $F_X(x) = P\{X \le x\}$. Như vậy khi ta biết qui luật phân bố xác suất của một biến ngẫu nhiên thì ta đã nắm được toàn bộ thông tin về biến ngẫu nhiên này.

Trường hợp biến ngẫu nhiên chỉ nhận các giá trị rời rạc thì hàm phân bố xác suất hoàn toàn được xác định bởi hàm khối lượng xác suất hoặc bảng phân bố xác suất, đó là bảng ghi các giá trị mà biến ngẫu nhiên nhận với xác suất tương ứng. Đối với biến ngẫu nhiên nhận giá trị liên tục thì hàm phân bố xác suất có thể được xác định bởi hàm mật độ xác suất.

Ngoài phương pháp sử dụng hàm phân bố xác suất để xác định biến ngẫu nhiên, trong nhiều trường hợp bài toán chỉ cần đòi hỏi khảo sát những đặc trưng cơ bản của biến ngẫu nhiên.

Các đặc trưng của biến ngẫu nhiên được chia thành hai loại sau:

- Các đặc trưng cho vị trí trung tâm của biến ngẫu nhiên như: Kỳ vọng, Trung vị, Mốt.

- Các đặc trưng cho độ phân tán của biến ngẫu nhiên như: Phương sai, Độ lệch chuẩn, Hệ số biến thiên, Hệ số bất đối xứng và Hệ số nhọn.

Trong các bài toán thực tế, kỳ vọng được sử dụng dưới dạng lợi nhuận kỳ vọng còn phương sai để tính mức độ rủi ro của quyết định. Trong kỹ thuật độ lệch chuẩn biểu diễn sai số của phép đo.

Một số quy luật phân bố xác suất quan trọng:

- Quy luật nhị thức, quy luật này thường gặp trong dãy phép thử Bernoulli.

- Quy luật Poisson, quy luật này thường gặp trong bài toán về quá trình đếm sự xuất hiện biến cố A nào đó. Quá trình đến trong các hệ phục vụ.

- Quy luật phân bố đều trong một khoảng là quy luật phân bố xác suất của biến ngẫu nhiên liên tục đồng khả năng lấy giá trị trong khoảng đó. Quy luật phân bố đều có ứng dụng rộng rãi trong thống kê toán. Nó có ý nghĩa to lớn trong các bài toán sử dụng phương pháp phi tham số.

- Quy luật mũ.

- Quy luật Erlang.

- Quy luật chuẩn.

- Quy luật "khi bình phương".

- Quy luật Student.

Phân bố chuẩn thường được gặp trong các bài toán về sai số khi đo đạc các đại lượng trong vật lý, thiên văn... Trong thực tế, nhiều biến ngẫu nhiên tuân theo quy luật chuẩn hoặc tiệm cận chuẩn (định lý giới hạn trung tâm), chẳng hạn: trọng lượng, chiều cao của một nhóm người nào đó, điểm thi của thí sinh, năng suất cây trồng, mức lãi suất

của một công ty, nhu cầu tiêu thụ của một mặt hàng nào đó... là các biến ngẫu nhiên có phân bố chuẩn.

Với mỗi quy luật phân bố xác suất ta sẽ khảo sát bảng phân bố xác suất hoặc hàm mật độ xác suất, các đặc trưng và các tính chất của nó.

Để học tốt chương này học viên phải nắm vững định nghĩa xác suất, biến cố và các tính chất của chúng.

Các đặc trưng của biến ngẫu nhiên được định nghĩa thông qua tính tổng của các số hạng nào đó (trường hợp biến ngẫu nhiên rời rạc) hoặc tính tích phân xác định, tích phân suy rộng với cận vô cùng (trường hợp biến ngẫu nhiên liên tục). Vì vậy bạn đọc cần ôn tập về chuỗi số, tích phân xác định và tích phân suy rộng.

2.1 ĐỊNH NGHĨA VÀ PHÂN LOẠI BIẾN NGẪU NHIÊN

Gieo một con xúc sắc 6 mặt. Ký hiệu A_1, A_2, A_3, A_4, A_5, A_6 lần lượt là biến cố "mặt 1 chấm xuất hiện", "mặt 2 chấm xuất hiện",…, "mặt 6 chấm xuất hiện".

Nếu xét đại lượng X là số chấm xuất hiện khi gieo con xúc sắc thì X có thể nhận các giá trị 1, 2, 3, 4, 5, 6 một cách ngẫu nhiên và X nhận giá trị k là biến cố A_k, nghĩa là $\{X = k\} = A_k$, với $k = 1,2,...,6$.

Ta gọi X là một biến ngẫu nhiên có miền giá trị $R_X = \{1,2,...,6\}$.

Một cách tổng quát ta có khái niệm biến ngẫu nhiên như sau:

2.1.1 Định nghĩa biến ngẫu nhiên

Định nghĩa 2.1: *Biến ngẫu nhiên X là đại lượng nhận các giá trị nào đó phụ thuộc vào các yếu tố ngẫu nhiên, nghĩa là với mọi giá trị thực $x \in \mathbb{R}$ thì "X nhận giá trị nhỏ hơn bằng x", ký hiệu $\{X \le x\}$, là một biến cố ngẫu nhiên.*

Đối với biến ngẫu nhiên người ta chỉ quan tâm xem nó nhận một giá trị nào đó hoặc nhận giá trị trong một khoảng nào đó với một xác suất bao nhiêu.

Tập hợp tất cả các giá trị của X được gọi là miền giá trị của X, ký hiệu R_X.

Ví dụ 2.1: Gieo đồng thời hai con xúc sắc. Ký hiệu A_k, $k = 2,3,...,12$ là biến cố tổng số chấm xuất hiện của hai con xúc sắc là k.

Nếu gọi X là tổng số chấm xuất hiện khi gieo hai con xúc sắc thì X là một biến ngẫu nhiên có miền giá trị $R_X = \{2,3,...,12\}$ và $\{X = k\} = A_k$ với $k = 2,3,...,12$.

Ví dụ 2.2: Các đại lượng sau là biến ngẫu nhiên

- Số chấm xuất hiện khi gieo một con xúc sắc.

- Tuổi thọ của một thiết bị đang hoạt động.

- Số khách hàng vào một điểm phục vụ trong một khoảng thời gian nào đó.

- Số cuộc gọi đến một tổng đài trong một khoảng thời gian nào đó.

- Sai số khi đo lường một đại lượng vật lý…

Định nghĩa 2.2: *Hai biến ngẫu nhiên X, Y là độc lập nếu X nhận các giá trị nào đó không phụ thuộc Y và ngược lại. Nói cách khác với mọi số thực x, y hai biến cố sau là độc lập:*

$$\{X \le x\},\ \{Y \le y\}$$

Trong chương 3 ta sẽ đưa ra các tiêu chuẩn để nhận biết tính chất độc lập của hai biến ngẫu nhiên.

2.1.2 Hàm phân bố xác suất

Các biến ngẫu nhiên xét trong các phép thử khác nhau (tương ứng với các không gian xác suất khác nhau) nhưng các quy luật phân bố xác suất của chúng có thể như nhau. Chẳng hạn xác suất bắn trúng bia của một xạ thủ là 0,8. Xạ thủ này bắn 10 viên, gọi X là số viên bắn trúng bia thì $P\{X = k\} = C_{10}^k (0,6)^k (0,4)^{10-k}$, $0 \le k \le 10$ (xem phân bố nhị thức, mục 2.2.2.2). Tương tự giả sử tỷ lệ chính phẩm của lô hàng

là 0,8. Chọn 10 sản phẩm kiểm tra, gọi Y là số chính phẩm phát hiện được thì $P\{Y=k\}=C_{10}^{k}(0,6)^{k}(0,4)^{10-k}$, $0\le k\le 10$. Nói cách khác quy luật phân bố xác suất của X và Y như nhau.

Phân bố xác suất được nghiên cứu thông qua hàm phân bố xác suất.

Định nghĩa 2.3: *Hàm phân bố xác suất (cumulative distribution function, viết tắt CDF) của biến ngẫu nhiên X là hàm số $F_X(x)$ xác định với mọi $x\in\mathbb{R}$ bởi công thức:*

$$F_X(x)=P\{X\le x\};\ -\infty<x<\infty \tag{2.1}$$

trong đó $\{X\le x\}$ là ký hiệu biến cố "biến ngẫu nhiên X nhận giá trị nhỏ hơn hay bằng x".

Hàm phân bố có các tính chất sau:

a) $0\le F_X(x)\le 1$ với mọi $x\in\mathbb{R}$, (2.2)

b) $F_X(x)$ là hàm không giảm, liên tục bên phải. Nếu X là biến ngẫu nhiên liên tục thì $F_X(x)$ là hàm liên tục.

$$F_X(a^+)=F_X(a) \text{ với } F_X(a^+)=\lim_{x>a,x\to a}F_X(x) \tag{2.3}$$

c) $F_X(-\infty)=\lim_{x\to-\infty}F_X(x)=0;\ F_X(+\infty)=\lim_{x\to+\infty}F_X(x)=1$ (2.4)

d) $P\{a<X\le b\}=F_X(b)-F_X(a)$ (2.5)

e) $P\{X>a\}=1-F_X(a)$; $P\{X<a\}=F_X(a^-)$ với

$$F_X(a^-)=\lim_{x<a,x\to a}F_X(x) \tag{2.6}$$

Nhận xét 2.1: Một số tài liệu coi $G_X(x)=P\{X<x\};\ -\infty<x<\infty$ là hàm phân bố của biến ngẫu nhiên X. Có một số khác biệt giữa hai định nghĩa này. Chẳng hạn tính chất liên tục phải của $F_X(x)$ được thay bằng liên tục trái của $G_X(x)$, công thức (2.5) sẽ là:

$$P\{a\le X<b\}=G_X(b)-G_X(a).$$

Ví dụ 2.3: Một nguồn tin sinh ra các ký hiệu ngẫu nhiên từ bốn ký tự $\{a,b,c,d\}$ với xác suất $P(a)=1/2$, $P(b)=1/4$ và $P(c)=P(d)=1/8$. Mã hóa các ký hiệu này theo các mã nhị phân sau

a	0
b	10
c	110
d	111

Đặt X là biến ngẫu nhiên ký hiệu độ dài của mã, đó là số các bit.

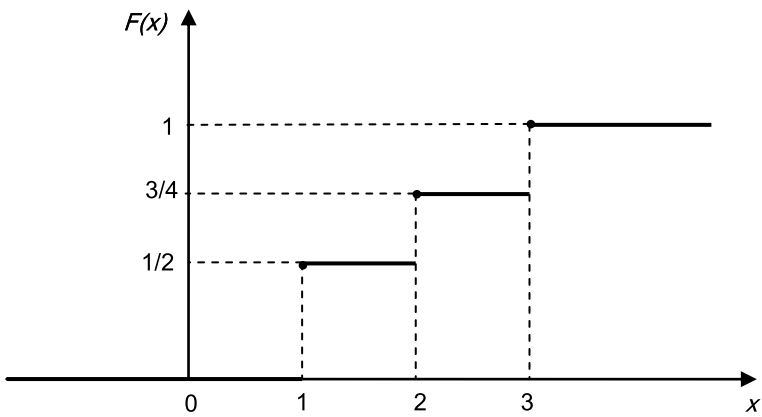
a) Tìm miền giá trị của X.

b) Giả sử các ký hiệu được sinh độc lập. Tính các xác suất

$$P\{X=1\},\ P\{X=2\} \text{ và } P\{X=3\}.$$

c) Tìm hàm phân bố $F_X(x)$ và vẽ đồ thị của $F_X(x)$.

Giải:



Hình 2.1

a) Miền giá trị $R_X=\{1,2,3\}$.

b) $P\{X=1\}=P(a)=1/2$, $P\{X=2\}=P(b)=1/4$,

$P\{X=3\}=P\{c,d\}=P(c)+P(d)=1/4$.

c) $F_X(x) = \begin{cases} 0 & x < 1 \\ 1/2 & 1 \le x < 2 \\ 3/4 & 2 \le x < 3 \\ 1 & x \ge 3 \end{cases}$

Đồ thị của $F_X(x)$ có dạng bậc thang (hình 2.1).

Ví dụ 2.4: Xét phép thử ném phi tiêu vào một đĩa tròn có bán kính bằng 1 (xem ví dụ 1.20). Ký hiệu X là biến ngẫu nhiên đo khoảng cách từ điểm mũi phi tiêu cắm vào đĩa đến tâm của đĩa. Giả sử mũi phi tiêu luôn cắm vào đĩa và đồng khả năng tại mọi điểm của đĩa.

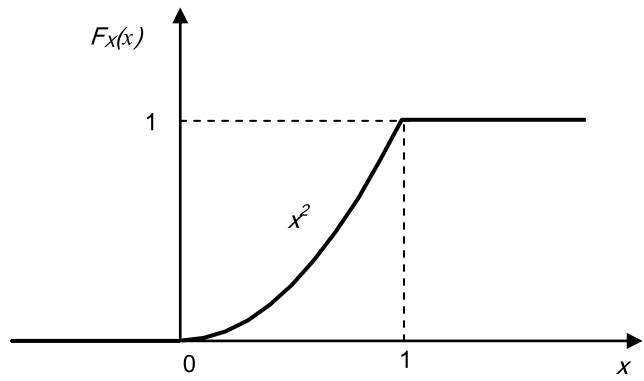
a) Tìm miền giá trị của X.

b) Tìm hàm phân bố $F_X(x)$ và vẽ đồ thị của $F_X(x)$.

Giải: a) Miền giá trị của X là $R_X = \{x \in 3 \mid 0 \le x < 1\}$.

b) $P\{X \le x\} = \dfrac{\pi.x^2}{\pi.1^2} = x^2$ nếu $0 \le x \le 1$.

Vậy hàm phân bố: $F_X(x) = \begin{cases} 0 & x \le 0 \\ x^2 & 0 < x \le 1 \\ 1 & x > 1 \end{cases}$



Hình 2.2

Ví dụ 2.5: Xét hàm số $F(x)=\begin{cases}0 & x\le 0\\ x+1 & 0<x\le 1/2\\ 1 & x>1/2\end{cases}$

a) Vẽ đồ thị của $F(x)$ và chứng minh rằng $F(x)$ thỏa mãn các tính chất (2.2); (2.4).

b) Nếu X là biến ngẫu nhiên có hàm phân bố $F(x)$. Tính:

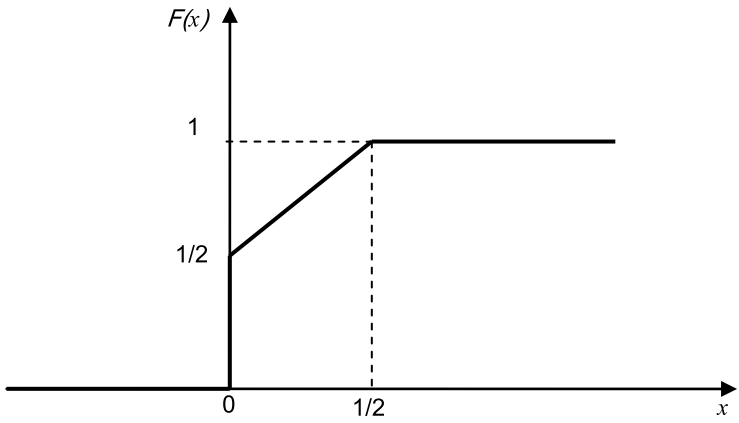
$P\{X\le 1/4\}$, $P\{0<X\le 1/4\}$, $P\{X=0\}$ và $P\{0\le X\le 1/4\}$.

Giải:

a) Từ đồ thị của $F(x)$ (hình 2.3) có thể suy ra các tính chất (2.2); (2.4) của hàm phân bố.

b) $P\left\{X\le\frac{1}{4}\right\}=F\left(\frac{1}{4}\right)=\frac{1}{4}+\frac{1}{2}=\frac{3}{4}$;

$$P\left\{0<X\le\frac{1}{4}\right\}=F\left(\frac{1}{4}\right)-F(0)=\frac{3}{4}-\frac{1}{2}=\frac{1}{4}$$



Hình 2.3

Áp dụng công thức (2.6) ta được:

$$P\{X=0\}=P\{X\le 0\}-P\{X<0\}=F(0)-F(0^-)=\frac{1}{2}-0=\frac{1}{2}$$

$$P\{0 \le X \le 1/4\} = P\{X = 0\} + P\{0 < X \le 1/4\} = \frac{1}{2} + \frac{3}{4} - \frac{1}{2} = \frac{3}{4}.$$

2.1.3 Phân loại

Hàm phân bố của biến ngẫu nhiên là hàm số không giảm và bị chặn do đó hàm phân bố chỉ có thể gián đoạn tại một số không quá đếm được các điểm, nghĩa là hoặc không có điểm gián đoạn hoặc nếu có điểm gián đoạn thì số điểm gián đoạn là đếm được. Dựa vào dạng hàm phân bố của biến ngẫu nhiên ta có thể chia biến ngẫu nhiên thành ba loại: Biến ngẫu nhiên rời rạc, biến ngẫu nhiên liên tục và loại hỗn hợp. Tuy nhiên chúng ta thường chỉ gặp hai loại biến ngẫu nhiên rời rạc hoặc liên tục.

- Biến ngẫu nhiên rời rạc nếu miền giá trị gồm một số hữu hạn hoặc vô hạn đếm được các giá trị, nghĩa là có thể liệt kê các giá trị thành một dãy $x_1, x_2, \ldots$. Do đó hàm phân bố có đồ thị dạng hình thang (hình 2.1, ví dụ 2.3)

- Biến ngẫu nhiên liên tục nếu miền giá trị của nó có thể lấp đầy một hoặc một số các khoảng hữu hạn hoặc vô hạn và xác suất biến ngẫu nhiên nhận giá trị tại từng điểm đều bằng 0 (nghĩa là $P\{X = a\} = 0$ với mọi a). Do đó hàm phân bố là hàm số liên tục (hình 2.2, ví dụ 2.4).

Biến ngẫu nhiên có hàm phân bố ở ví dụ 2.4 thuộc loại hỗn hợp, không rời rạc và không liên tục.

Ví dụ 2.6:

- Gọi X là số chấm xuất hiện khi gieo một con xúc sắc thì X là biến ngẫu nhiên rời rạc nhận các giá trị $1, 2, 3, 4, 5, 6$.

- Gọi Y là tuổi thọ của một thiết bị đang hoạt động thì Y là biến ngẫu nhiên liên tục nhận giá trị trong một khoảng.

- Gọi Z là số khách hàng vào một điểm phục vụ trong 1 đơn vị thời gian, Z là biến ngẫu nhiên rời rạc nhận các giá trị $0, 1, 2, \ldots$

- Số cuộc gọi đến một tổng đài là biến ngẫu nhiên rời rạc nhận các giá trị $0,1,2,...$

- Sai số Y khi đo lường một đại lượng vật lý nào đó là biến ngẫu nhiên liên tục nhận giá trị trong một khoảng.

2.2 BIẾN NGẪU NHIÊN RỜI RẠC

2.2.1 Hàm khối lượng xác suất và bảng phân bố xác suất của biến ngẫu nhiên rời rạc

Biến ngẫu nhiên rời rạc có miền giá trị là một tập hữu hạn hoặc vô hạn đếm được. Các xác suất chỉ tập trung tại các giá trị này.

Giả sử biến ngẫu nhiên rời rạc X có đồ thị của $F_X(x)$ là hàm bậc thang có bước nhảy tại x_1, x_2, ... thì:

$$F_X(x_k) - F_X(x_{k-1}) = P\{X \le x_k\} - P\{X \le x_{k-1}\} = P\{X = x_k\} \quad (2.7)$$

Đặt $$p_X(x) = P\{X = x\} \quad (2.8)$$

Hàm $p_X(x)$ được *gọi là hàm khối lượng xác suất* (probability mass function) *của biến ngẫu nhiên rời rạc* X.

Tính chất của hàm khối lượng xác suất $p_X(x)$:

1. $p_X(x_k) > 0$ (2.9)

2. $\sum_k p_X(x_k) = 1$ (2.10)

3. $p_X(x) = 0$ với mọi $x \ne x_k$ (2.11)

Hàm phân bố của X được tính theo công thức:

$$F_X(x) = P\{X \le x\} = \sum_{x_k \le x} p_X(x_k) \quad (2.12)$$

- Nếu biến ngẫu nhiên rời rạc X nhận vô hạn các giá trị x_1, x_2,... thì hàm phân bố xác suất có dạng:

$$F_X(x)=\begin{cases} 0 & \text{nếu} \quad x<x_1 \\ p_X(x_1)+\cdots+p_X(x_{k-1}) & \text{nếu} \quad x_{k-1}\le x<x_k\,,\ \forall k>1 \end{cases} \tag{2.13}$$

- Nếu X chỉ nhận một số hữu hạn các giá trị $x_1, x_2, ..., x_n$ thì các biến cố:

$$\{X=x_1\}, \{X=x_2\}, ..., \{X=x_n\} \tag{2.14}$$

lập thành hệ đầy đủ các biến cố.

Hàm phân bố xác suất có dạng:

$$F_X(x)=\begin{cases} 0 & \text{nếu} \quad x<x_1 \\ p_X(x_1)+\cdots+p_X(x_{k-1}) & \text{nếu} \quad x_{k-1}\le x<x_k \\ 1 & \text{nếu} \quad x\ge x_n \end{cases} \tag{2.15}$$

Chúng ta cũng có thể biểu diễn phân bố của biến ngẫu nhiên rời rạc thông qua *bảng phân bố xác suất*. Đó là bảng có hai hàng, hàng trên ghi các giá trị mà biến ngẫu nhiên nhận được, hàng dưới là giá trị của hàm khối lượng xác suất tương ứng.

Bảng phân bố xác suất của X có dạng sau:

X	x_1	x_2	...
P	$p_X(x_1)$	$p_X(x_2)$	...

Ví dụ 2.7: Xét phép thử tung đồng thời 2 đồng xu (Ví dụ 1.1). Không gian mẫu của phép thử là $\Omega=\{(S,S),(S,N),(N,S),(N,N)\}$ gồm 4 kết quả đồng khả năng. Gọi X là số mặt sấp xuất hiện, khi đó X là một biến ngẫu nhiên rời rạc.

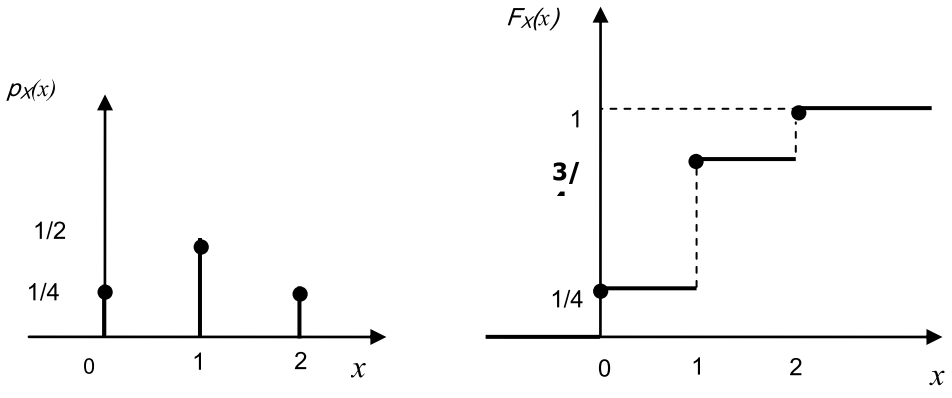
Bảng phân bố xác suất:

X	0	1	2
P	1/4	2/4	1/4

Hàm phân bố xác suất

$$F_X(x)=\begin{cases} 0 & \text{nếu} \quad x<0 \\ 1/4 & \text{nếu} \quad 0\le x<1 \\ 3/4 & \text{nếu} \quad 1\le x<2 \\ 1 & \text{nếu} \quad x\ge 2 \end{cases}$$

Đồ thị của hàm khối lượng xác suất và hàm phân bố



(*a*) (*b*)

Hình 2.4

Ví dụ 2.8: Chọn ngẫu nhiên 3 bi từ một túi có 6 bi đen, 4 bi trắng. Gọi X là số bi trắng trong 3 bi vừa chọn thì X là một biến ngẫu nhiên rời rạc. Tìm bảng phân bố xác suất và hàm phân bố xác suất.

Giải:

$$P\{X=0\}=\frac{C_6^3}{C_{10}^3}=\frac{5}{30},\quad P\{X=1\}=\frac{C_6^2C_4^1}{C_{10}^3}=\frac{15}{30},$$

$$P\{X=2\}=\frac{C_6^1C_4^2}{C_{10}^3}=\frac{9}{30},\quad P\{X=3\}=\frac{C_4^3}{C_{10}^3}=\frac{1}{30}$$

Hàm khối lượng xác suất

$$p_X(0)=\frac{5}{30},\ p_X(1)=\frac{15}{30},\ p_X(2)=\frac{9}{30},\ p_X(3)=\frac{1}{30};$$

$p_X(x)=0$ với mọi $x \neq 0, 1, 2, 3$.

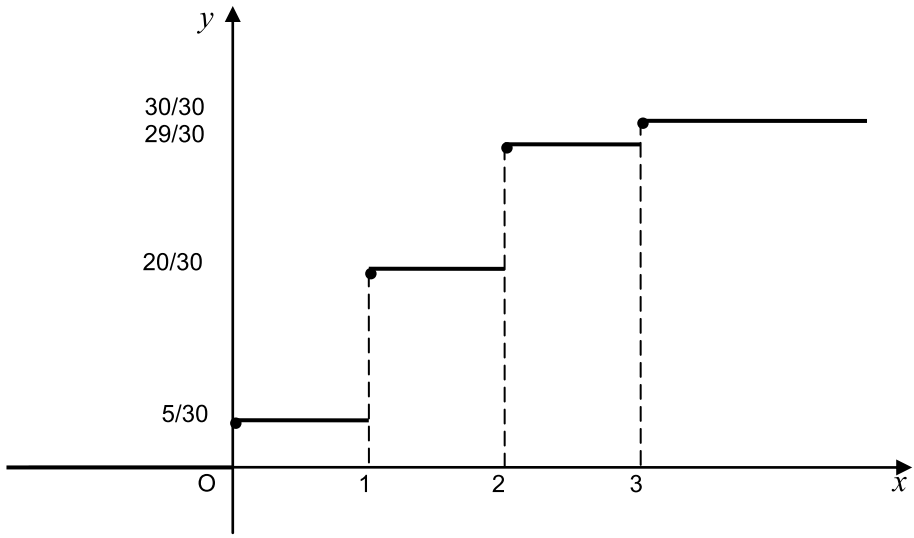
Bảng phân bố xác suất:

X	0	1	2	3
P	5/30	15/30	9/30	1/30

Hàm phân bố xác suất:

$$F_X(x)=\begin{cases} 0 & \text{nếu} \quad x<0 \\ 5/30 & \text{nếu} \quad 0\le x<1 \\ 20/30 & \text{nếu} \quad 1\le x<2 \\ 29/30 & \text{nếu} \quad 2\le x<3 \\ 1 & \text{nếu} \quad x\ge 3 \end{cases}$$

Đồ thị hình 2.5:



Hình 2.5

Ví dụ 2.9: Xác suất để xạ thủ bắn trúng bia là p, $0<p<1$. Xạ thủ được phát từng viên đạn để bắn cho đến khi trúng bia. Hãy xây dựng bảng phân bố xác suất của số viên đạn được phát.

Giải: Gọi X là "số viên đạn được phát". X là biến ngẫu nhiên rời rạc với các giá trị có thể có là $1,2,...,k,...$ Gọi A_k là biến cố "xạ thủ bắn trúng bia ở viên đạn thứ k". Các biến cố A_k độc lập nhau và $P(A_k)=p$.

$$P\{X=1\}=P(A_1)=p$$

$$P\{X=2\}=P\left(\overline{A_1}A_2\right)=P\left(\overline{A_1}\right)P\left(A_2\right)=qp,\ q=1-p$$

Tổng quát:

$$P\{X=k\}=P\left(\overline{A_1}...\overline{A_{k-1}}A_k\right)=P\left(\overline{A_1}\right)...P\left(\overline{A_{k-1}}\right)P\left(A_k\right)=q^{k-1}p$$

Hàm khối lượng xác suất

$$p_X(x)=\begin{cases}q^{k-1}p & x=k\\ 0 & x\neq k\end{cases};\ k=1,2,...;\ 0<p<1 \qquad (2.16)$$

Bảng phân bố xác suất có dạng:

X	1	2	...	k	...
P	p	qp	...	$q^{k-1}p$	...

Biến ngẫu nhiên với phân bố rời rạc có dạng trên được gọi là *biến ngẫu nhiên cấp số nhân* (geometric random variable).

2.2.2 Các phân bố rời rạc thường gặp

2.2.2.1 Phân bố Bernoulli

Định nghĩa 2.4: *Biến ngẫu nhiên rời rạc X nhận hai giá trị $0, 1$ với xác suất tương ứng*

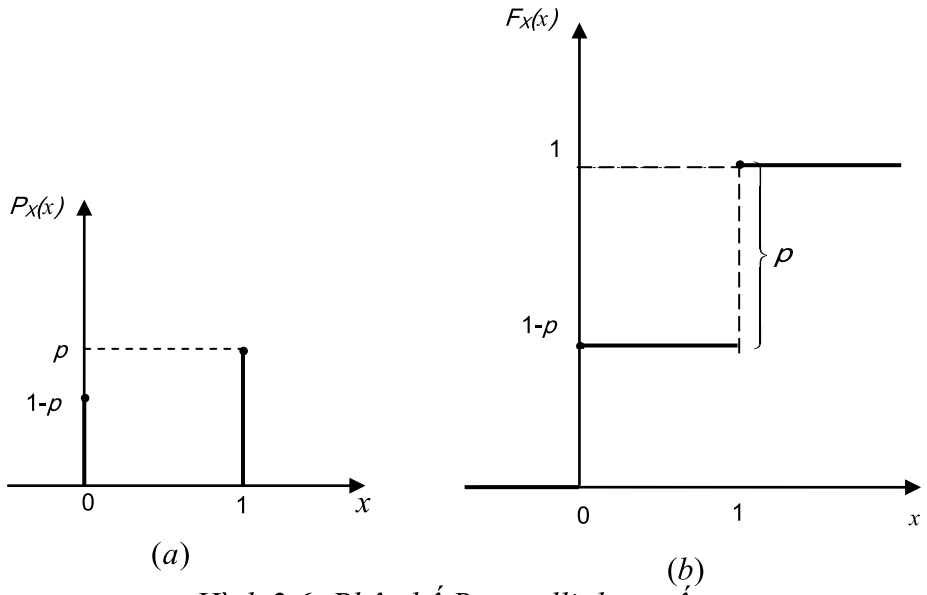
$$p_X(k)=P\{X=k\}=p^k q^{1-k};\ k=0,1 \qquad (2.17)$$

trong đó $0<p<1, q=1-p$, được gọi là có phân bố Bernoulli tham số p.

Xét phép thử Bernoulli với sự thành công của phép thử là sự xuất hiện của biến cố A và giả sử xác suất xuất hiện của A trong mỗi lần thử là p. Gọi X là số lần thành công trong một lần thử thì X là biến ngẫu nhiên rời rạc có phân bố Bernoulli tham số p.

Phân bố Bernoulli tham số p còn được gọi là phân bố không – một $A(p)$.

Hình 2.6 minh họa phân bố Bernoulli tham số p.



Hình 2.6: Phân bố Bernoulli tham số p

2.2.2.2 Phân bố nhị thức $\mathcal{B}(n;p)$

Định nghĩa 2.5: *Biến ngẫu nhiên rời rạc X nhận các giá trị* 0, 1, ..., *n với xác suất tương ứng:*

$$p_X(k) = P\{X = k\} = C_n^k p^k q^{n-k};\ k = 0,1,...,n \qquad (2.18)$$

trong đó n là số tự nhiên và $0 < p < 1$, $q = 1 - p$, *được gọi là có phân bố nhị thức tham số n, p, ký hiệu* $X \sim \mathcal{B}(n;p)$.

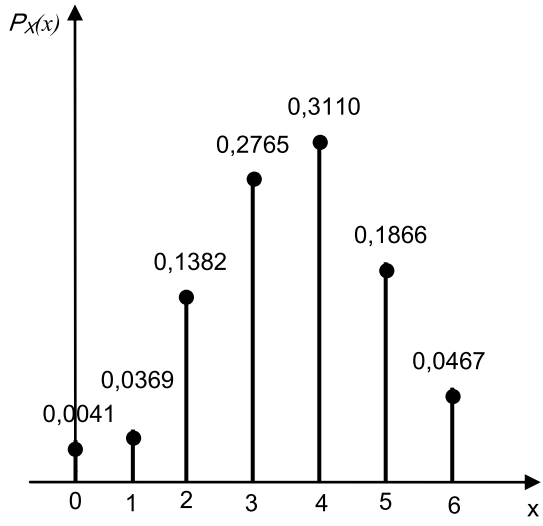
Bảng phân bố xác suất của biến ngẫu nhiên có quy luật nhị thức $\mathcal{B}(n;p)$

X	0	1	...	*k*	...	*n*
P	$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$	...	$C_n^k p^k q^{n-k}$	...	$C_n^n p^n q^0$

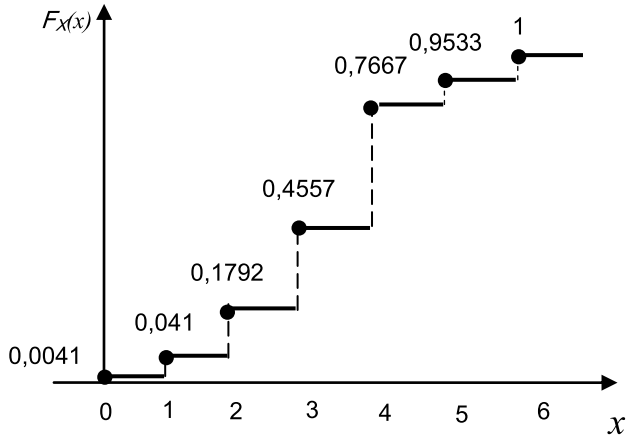
Hàm phân bố

$$F_X(x) = \sum_{k=0}^{m} C_n^k p^k q^{n-k}, \quad m \le x < m+1; \; m \le n \tag{2.19}$$

Hình 2.7 minh họa phân bố nhị thức với $n = 6$ và $p = 0,6$.



a)



b)

Hình 2.7: Phân bố nhị thức với n = 6 và p = 0,6

Nhận xét 2.2:

1. Thực hiện n phép thử Bernoulli với xác suất thành công của biến cố A trong mỗi lần thử là p.

Với mỗi $i = 1,2,...,n$; nếu ở lần thử thứ i biến cố A xuất hiện ta cho X_i nhận giá trị 1, nếu biến cố A không xuất hiện ta cho X_i nhận giá trị 0. Như vậy X_i là biến ngẫu nhiên rời rạc có phân bố Bernoulli tham số p:

Gọi X là số thành công trong n phép thử Bernoulli này thì:

$$X = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n; p) \tag{2.20}$$

2. Từ (2.20) suy ra rằng nếu $X \sim \mathcal{B}(n_1; p)$, $Y \sim \mathcal{B}(n_2; p)$ và X, Y độc lập thì:

$$X + Y \sim \mathcal{B}(n_1 + n_2; p) \tag{2.21}$$

Ví dụ 2.10: Tỉ lệ phế phẩm của lô hàng là 4%. Chọn ngẫu nhiên 20 sản phẩm để kiểm tra. Gọi X là số phế phẩm phát hiện được.

a) Gọi tên luật phân bố xác suất của X.

b) Tính xác suất có đúng 5 phế phẩm phát hiện được.

c) Lô hàng được xem là đạt tiêu chuẩn nếu số phế phẩm phát hiện được không nhiều hơn 2. Tính xác suất để lô hàng đạt tiêu chuẩn.

Giải: Có thể xem kiểm tra chất lượng mỗi sản phẩm là thực hiện một phép thử Bernoulli với sự thành công của phép thử là phát hiện ra phế phẩm. Theo giả thiết xác suất thành công của mỗi lần thử là 0,04. Kiểm tra 20 sản phẩm là thực hiện 20 phép thử.

a) Số phế phẩm phát hiện được là số lần thành công trong 20 phép thử này. Vậy X có phân bố nhị thức $X \sim \mathcal{B}(n; p)$, với $n = 20,\ p = 0,04$.

b) Xác suất phát hiện đúng 5 phế phẩm là:

$$P\{X = 5\} = C_{20}^{5} (0,04)^5 (0,96)^{15} = 0,0008$$

Xác suất để lô hàng đạt tiêu chuẩn là $P\{X \le 2\} = 0,956$

Ví dụ 2.11: Một nguồn nhị phân phát ra hai ký số (digit) 1 và 0 một cách ngẫu nhiên với xác suất tương ứng 0,6 và 0,4.

a) Tính xác suất có đúng hai ký số 1 và ba ký số 0 trong dãy có năm ký số.

b) Tính xác suất có ít nhất ba ký số 1 trong dãy có năm ký số.

Giải: Gọi X là số các ký số 1 trong dãy có năm ký số. Theo giả thiết chỉ có hai khả năng đầu vào có thể là 1 và 0, xác suất xuất hiện của 1 không đổi và bằng 0,6. Vậy có thể xem mỗi ký số xuất hiện trong dãy tương ứng với một phép thử Bernoulli mà sự thành công của mỗi lần thử là xuất hiện ký số 1. Mỗi dãy có năm ký số tương ứng với 5 phép thử Bernoulli, vậy X có phân bố nhị thức $X \sim \mathcal{B}(n;p)$, với $n=5,\ p=0,6$.

a) Xác suất để dãy có đúng hai ký số 1 và ba ký số 0 trong dãy có năm ký số là:

$$P\{X=2\} = C_5^2 (0,6)^2 (0,4)^3 = 0,23$$

b) Xác suất có ít nhất ba ký số 1 trong dãy có năm ký số là:

$$P\{X \geq 3\} = 1 - P\{X \leq 2\}$$

Trong đó: $P\{X \leq 2\} = \sum_{k=0}^{2} C_5^k (0,6)^k (0,4)^{5-k} = 0,317$

Vậy $P\{X \geq 3\} = 1 - 0,317 = 0,683$

2.2.2.3 Phân bố Poisson

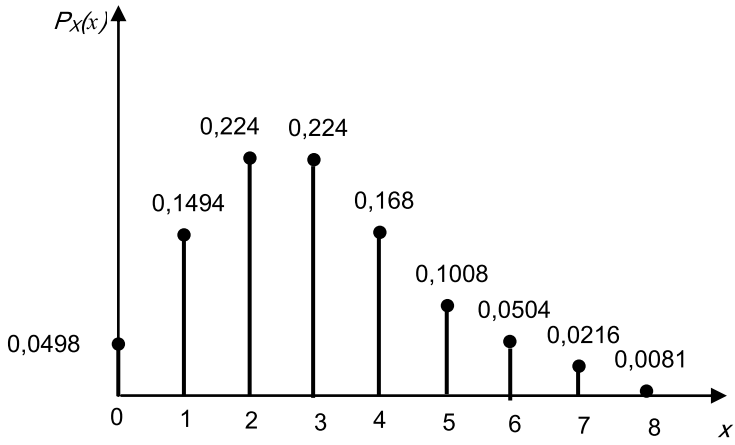
Định nghĩa 2.6: *Biến ngẫu nhiên X nhận các giá trị $k = 0, 1, 2, \ldots$ với hàm khối lượng xác suất*

$$p_X(k) = P\{X=k\} = e^{-\lambda} \frac{\lambda^k}{k!};\ k = 0,1,2,\ldots;\lambda > 0 \qquad (2.22)$$

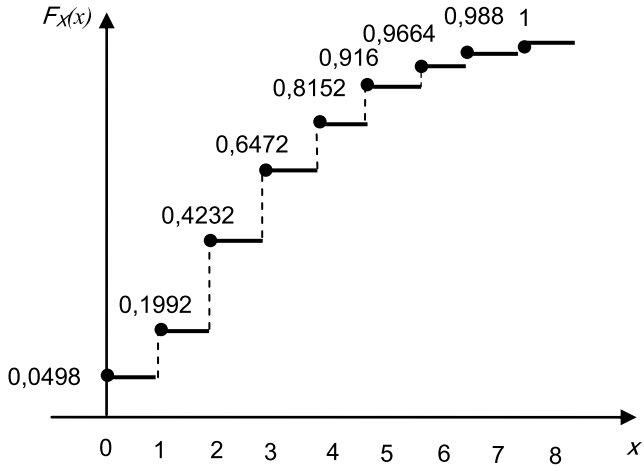
gọi là có phân bố Poisson tham số $\lambda > 0$, ký hiệu $X \sim \mathcal{P}(\lambda)$.

Hàm phân bố

$$F_X(x) = e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!}, \quad n \le x < n+1 \tag{2.23}$$



a)



b)

Hình 2.8: Phân bố Poisson với tham số $\lambda = 3$

Trong thực tế với một số giả thiết thích hợp thì các biến ngẫu nhiên là các quá trình đếm sau:

1) Số cuộc gọi đến một tổng đài,

2) Số khách hàng đến 1 điểm phục vụ,

3) Số xe cộ qua 1 ngã tư,

4) Số tai nạn (xe cộ); số các sự cố xảy ra ở một địa điểm,…

trong một khoảng thời gian xác định nào đó sẽ có phân bố Poisson với tham số λ, trong đó λ là tốc độ trung bình diễn ra trong khoảng thời gian này.

Hình 2.8 minh họa phân bố Poisson với tham số $\lambda = 3$.

Ví dụ 2.12: Ở một tổng đài điện thoại các cuộc gọi đến một cách ngẫu nhiên, độc lập và trung bình có 2 cuộc gọi trong 1 phút. Tìm xác suất để:

a) Có đúng 5 cuộc gọi đến trong 2 phút (biến cố A).

b) Không có một cuộc gọi nào trong 30 giây (biến cố B).

c) Có ít nhất 1 cuộc gọi trong 10 giây (biến cố C).

Giải: Ký hiệu $X(t)$ là số cuộc gọi đến tổng đài trong khoảng thời gian t phút, có thể chứng minh được $X(t)$ có phân bố Poisson tham số λt, trong đó λ là số cuộc gọi trung bình trong 1 phút. Theo giả thiết $\lambda = 2$.

a) $X(2) \sim \mathcal{P}(4)$, do đó: $P(A) = P\{X(2) = 5\} = e^{-4}\dfrac{4^5}{5!} \approx 0{,}156$

b) $X(1/2) \sim \mathcal{P}(1)$, do đó: $P(B) = P\{X(1/2) = 0\} = e^{-1} \approx 0{,}3679$

c) $X(1/6) \sim \mathcal{P}(1/3)$, do đó:

$$P(C) = P\{X(1/6) \geq 1\} = 1 - P\{X(1/6) = 0\} = 1 - e^{-1/3} \approx 0{,}2835$$

Quy luật Poisson có ứng dụng rộng rãi trong nhiều lĩnh vực thực tế như kiểm tra chất lượng sản phẩm, lý thuyết sắp hàng, các hệ phục vụ đám đông, các bài toán chuyển mạch trong tổng đài…

Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân bố Poisson tham số lần lượt λ_1, λ_2 thì $X_1 + X_2$ cũng có phân bố Poisson tham số $\lambda_1 + \lambda_2$ (ví dụ 3.10 chương 3).

$$X_1 + X_2 \sim \mathcal{P}(\lambda_1 + \lambda_2) \tag{2.24}$$

2.3 BIẾN NGẪU NHIÊN LIÊN TỤC

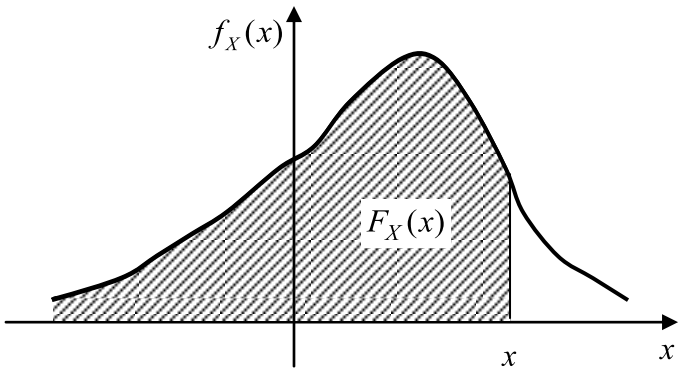
2.3.1 Hàm mật độ xác suất của biến ngẫu nhiên liên tục

Công thức (2.12) cho thấy hàm phân bố xác suất của biến ngẫu nhiên rời rạc có thể được xác định qua hàm khối lượng xác suất. Mục này sẽ chỉ ra rằng hàm phân bố xác suất của biến ngẫu nhiên liên tục có thể được xác định bởi hàm mật độ xác suất tương ứng.

Định nghĩa 2.7: *Giả sử X là một biến ngẫu nhiên liên tục có hàm phân bố xác suất $F_X(x)$, nếu tồn tại hàm $f_X(x)$ sao cho:*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \text{, với mọi } x \in \mathbb{R} \tag{2.25}$$

thì $f_X(x)$ được gọi là hàm mật độ xác suất của biến ngẫu nhiên X (probability density function, viết tắt PDF).



Hình 2.9

Như vậy giá trị của hàm $F_X(x)$ bằng diện tích hình phẳng giới hạn bởi đồ thị hàm mật độ xác suất $f_X(x)$, trục hoành và đường thẳng song song với trục tung có hoàng độ là x.

Tính chất của hàm mật độ xác suất $f_X(x)$

1. $F_X'(x) = f_X(x)$ tại các điểm x mà $f_X(x)$ liên tục. (2.26)

2. $f_X(x) \geq 0$ với mọi $x \in \mathbb{R}$. (2.27)

3. $\int_{-\infty}^{\infty} f_X(x)dx = 1$. (2.28)

4. $P\{a < X < b\} = P\{a \leq X \leq b\}$ (2.29)

$$= P\{a \leq X < b\} = P\{a < X \leq b\} = \int_a^b f_X(x)dx.$$

Ví dụ 2.13: Hàm phân bố xác suất của biến ngẫu nhiên liên tục X có dạng:

$$F_X(x) = \begin{cases} 0 & \text{với} \quad x < 0 \\ kx^2 & \text{với} \quad 0 \leq x < 1 \\ 1 & \text{với} \quad x \geq 1 \end{cases}$$

a) Xác định hệ số k;

b) Tìm hàm mật độ xác suất $f_X(x)$.

Giải:

a) Vì hàm phân bố xác suất $F_X(x)$ liên tục. Xét tại $x = 1$

$$\begin{cases} \lim\limits_{x \to 1^-} F_X(x) = \lim\limits_{x \to 1^-} kx^2 = k \\ F_X(1) = 1 \end{cases} \Rightarrow k = 1.$$

b) Theo tính chất (2.26) của hàm mật độ xác suất ta có:

$$f_X(x)=\begin{cases}0 & \text{với} \quad x\le 0\\ 2x & \text{với} \quad 0<x<1\\ 0 & \text{với} \quad x\ge 1\end{cases}$$

Ví dụ 2.14: Biến ngẫu nhiên liên tục X với hàm mật độ xác suất có dạng:

$$f_X(x)=\begin{cases}0 & \text{với} \quad x<1\\ \dfrac{k}{x^2} & \text{với} \quad x\ge 1\end{cases}$$

Hãy xác định:

a) Hệ số k;

b) Hàm phân bố xác suất $F_X(x)$;

c) Xác suất $P\{2<X<3\}$;

d) Xác suất để trong 4 phép thử độc lập biến ngẫu nhiên X đều không lấy giá trị trong khoảng $(2,3)$.

Giải:

a) Theo tính chất (2.28) ta có:

$$1=\int\limits_{-\infty}^{\infty} f_X(x)dx=\int\limits_{1}^{\infty}\frac{k}{x^2}dx=-\lim_{a\to\infty}\left(\frac{k}{x}\Big|_1^a\right)=k\text{, từ đó } k=1.$$

b) Từ công thức (2.25) ta có:

$$F_X(x)=\int\limits_{-\infty}^{x} f_X(t)dt=\begin{cases}0 & \text{với} \quad x<1\\ 1-\dfrac{1}{x} & \text{với} \quad x\ge 1\end{cases}$$

c) Từ công thức (2.29) và (2.5) ta có:

$$P\{2<X<3\}=F_X(3)-F_X(2)=\left(1-\frac{1}{3}\right)-\left(1-\frac{1}{2}\right)=\frac{1}{2}-\frac{1}{3}=\frac{1}{6}.$$

d) Theo kết quả ở trên xác suất để X không lấy giá trị trong khoảng $(2;3)$ trong một phép thử bằng $1-\frac{1}{6}=\frac{5}{6}$.

Vậy xác suất để trong 4 phép thử độc lập biến ngẫu nhiên X đều không lấy giá trị trong khoảng $(2;3)$ bằng $\left(\frac{5}{6}\right)^4 \approx 0,48$.

2.3.2 Các phân bố liên tục thường gặp

2.3.2.1 Phân bố đều U(a,b)

Định nghĩa 2.8: *Biến ngẫu nhiên X được gọi là có phân bố đều trong khoảng $(a;b)$ nếu hàm mật độ xác suất của nó xác định bởi:*

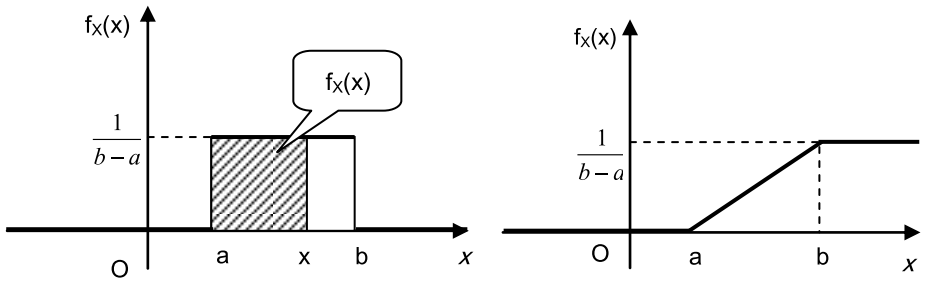
$$f_X(x)=\begin{cases}\dfrac{1}{b-a} & \text{nếu } a<x<b\\ 0 & \text{nếu ngược lại}\end{cases} \tag{2.30}$$

Hàm phân bố xác suất tương ứng

$$F_X(x)=\int_{-\infty}^{x} f_X(t)dt=\begin{cases}0 & \text{nếu } x\le a\\ \dfrac{x-a}{b-a} & \text{nếu } a<x<b\\ 1 & \text{nếu } x\ge b\end{cases} \tag{2.31}$$

Vậy X có khả năng nhận giá trị trong khoảng $(a;b)$ là "đều nhau" và không nhận giá trị ngoài $(a;b)$.

Quy luật phân bố đều có nhiều ứng dụng trong thống kê toán như mô phỏng thống kê, đặc biệt trong phương pháp phi tham số. Trong một số lý thuyết kết luận thống kê người ta thường xuất phát từ quy tắc sau đây: Nếu ta không biết gì về giá trị của tham số cần ước lượng thì mỗi giá trị có thể có của tham số đó là đồng khả năng, điều đó dẫn đến việc quan niệm tham số cần ước lượng như một biến ngẫu nhiên có quy luật phân bố đều.



Hình 2.10: Phân bố đều $\mathbf{U}(a,b)$

2.3.2.2 Phân bố mũ

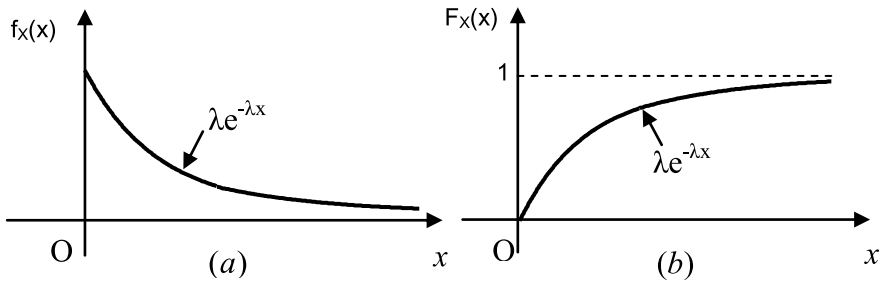
Định nghĩa 2.9: *Biến ngẫu nhiên liên tục* X *được gọi là có phân bố mũ tham số* $\lambda > 0$ *nếu hàm mật độ xác suất xác định như sau:*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \le 0 \end{cases} \tag{2.32}$$

Hàm phân bố xác suất

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 1 - e^{-\lambda x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \le 0 \end{cases} \tag{2.33}$$

Phân bố mũ thường xuất hiện trong các bài toán về thời gian sống của một loài sinh vật, tuổi thọ của thiết bị... hoặc khoảng thời gian giữa hai lần xuất hiện của một biến cố E nào đó mà số lần xuất hiện của E tuân theo luật phân bố Poisson.



Hình 2.11: Phân bố mũ tham số $\lambda = 1/7$

Ví dụ 2.15: Tuổi thọ của một mạch điện tử trong máy tính là một biến ngẫu nhiên có phân bố mũ tham số $\lambda > 0$. Giả sử tuổi thọ trung bình của mạch điện tử này là $\frac{1}{\lambda} = 6,25$ (năm) (xem kỳ vọng của phân bố mũ ở mục 2.5.3). Thời gian bảo hành là 2 năm. Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Giải: Gọi X là tuổi thọ của mạch điện tử. Xác suất để mạch điện tử bị hỏng trong thời gian bảo hành là:

$$P\{X \le 2\} = 1 - e^{-2\lambda} = 1 - e^{\frac{-2}{6,25}} = 1 - e^{-0,32} = 1 - 0,726 = 0,274 .$$

Vậy có khoảng 27,4% số mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Tính chất thú vị nhất của phân bố mũ là tính "không nhớ".

Biến ngẫu nhiên X được gọi là "không nhớ" (memoryless) nếu

$$P\{X > x + t | X > t\} = P\{X > x\} \quad \forall x, t > 0$$

Điều này tương đương với

$$P\{X > x + t\} = P\{X > x\}P\{X > t\} \quad \forall x, t > 0 \qquad (2.34)$$

Gọi $F(x)$ là hàm phân bố xác suất của X, đặt:

$$G(x) = P\{X > x\} = 1 - F(x).$$

Điều kiện (2.34) có thể viết lại:

$$G(x + t) = G(x)G(t) \qquad (2.35)$$

Giải phương trình (2.35) (phương trình hàm Cauchy) với điều kiện $G(x) = 1, \forall x < 0$ và $G(+\infty) = 0$ ta được $G(x) = e^{-\lambda x}$.

Vậy biến ngẫu nhiên X "không nhớ" khi và chỉ khi X có phân bố mũ. Chính vì thế phân bố mũ còn được gọi là *phân bố Markov*.

Ví dụ 2.16: Giả sử thời gian mỗi cuộc gọi điện thoại (tính theo phút) là một biến ngẫu nhiên X với phân bố mũ tham số $\lambda = 1/10$.

Một bốt điện thoại chỉ phục vụ từng người và giả sử A vào bốt điện thoại trước khi B đến. Tính xác suất B phải chờ đến lượt mình trong khoảng thời gian:

a) Ít hơn 5 phút.

b) Trong khoảng từ 5 đến 10 phút.

Giải: a) Vì tính chất "không nhớ" của phân bố mũ do đó thời gian chờ của B bằng thời gian A tiếp tục hoàn thành cuộc gọi tính từ lúc B đến và không phụ thuộc A đã gọi trong thời gian bao lâu. Theo công thức (2.33) ta được xác suất B phải chờ ít hơn 5 phút là:

$$P\{X<5\} = F_X(5) = 1-e^{-0,5} \approx 0,393\,.$$

b) Lập luận tương tự trên ta được xác suất B phải chờ trong khoảng từ 5 đến 10 phút là:

$$P\{5<X<10\} = F_X(10) - F_X(5)$$
$$= (1-e^{-1}) - (1-e^{-0,5}) = e^{-0,5} - e^{-1} \approx 0,239.$$

2.3.2.3 Phân bố Erlang

Định nghĩa 2.10: *Biến ngẫu nhiên X có phân bố Erlang tham số $(k;\lambda)$; k là số tự nhiên và $\lambda > 0$, nếu hàm mật độ xác suất có dạng:*

$$f_X(x) = \begin{cases} \dfrac{\lambda^k}{(k-1)!} x^{k-1} e^{-\lambda x} & \text{nếu} \quad x > 0 \\ 0 & \text{nếu} \quad x \le 0 \end{cases} \tag{2.36}$$

Có thể chứng minh được rằng nếu $X_1, X_2, ..., X_k$ là k biến ngẫu nhiên độc lập cùng có phân bố mũ tham số $\lambda > 0$ thì:

$$X = X_1 + X_2 + \cdots + X_k$$

có phân bố Erlang tham số $(k;\lambda)$.

2.3.2.4 Phân bố chuẩn $\mathbf{N(\mu;\sigma^2)}$

2.3.2.4.1 Định nghĩa

Định nghĩa 2.11: *Biến ngẫu nhiên liên tục X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, ký hiệu $X \sim \mathbf{N}(\mu;\sigma^2)$, nếu hàm mật độ xác suất có dạng:*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}; \quad \forall x \in 3 \tag{2.37}$$

Phân bố chuẩn được Gauss tìm ra năm 1809 nên nó còn được gọi là phân bố Gauss. Phân bố chuẩn thường được thấy trong các bài toán về sai số gặp phải khi đo đạc các đại lượng vật lý, thiên văn ...

Trong thực tế, có nhiều biến ngẫu nhiên tuân theo quy luật chuẩn hoặc tiệm cận chuẩn (Định lý giới hạn trung tâm). Chẳng hạn: trọng lượng, chiều cao của một nhóm người nào đó, điểm thi của thí sinh, năng suất cây trồng, mức lãi suất của một công ty, nhu cầu tiêu thụ của một mặt hàng nào đó, nhiễu trắng trên các kênh thông tin... là biến ngẫu nhiên có phân bố chuẩn.

2.3.2.4.2 Tính chất đồ thị của hàm mật độ xác suất của quy luật chuẩn

Từ công thức xác định hàm mật độ xác suất (2.26) ta suy ra các tính chất sau của đồ thị:

- Nhận trục $x = \mu$ làm trục đối xứng.

- Tiệm cận với trục hoành khi $x \to \pm\infty$.

- Diện tích giới hạn bởi đồ thị và trục hoành bằng 1.

- Đạt cực đại tại $x = \mu$ và có giá trị cực đại bằng $\frac{1}{\sigma\sqrt{2\pi}}$.

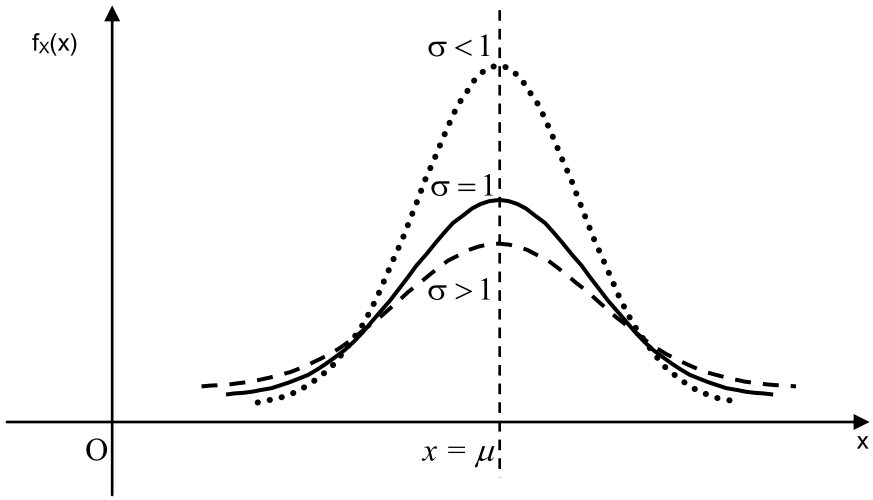
- Có 2 điểm uốn tại $x = \mu \pm \sigma$.

- Do đó khi μ tăng lên thì đồ thị dịch sang phải, còn khi μ giảm đồ thị dịch sang trái.

- Khi σ tăng lên thì đồ thị sẽ thấp xuống, còn khi σ giảm đồ thị cao lên và nhọn hơn.

Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân bố chuẩn $X_1 \sim \mathbf{N}(\mu_1; \sigma_1^2)$ và $X_2 \sim \mathbf{N}(\mu_2; \sigma_2^2)$ thì tổ hợp tuyến tính bất kỳ của X_1, X_2 cũng có phân bố chuẩn, đặc biệt:

$$X_1 + X_2 \sim \mathbf{N}(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2) \tag{2.38}$$



Hình 2.12: Đồ thị hàm mật độ của phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$

2.3.2.4.3 Phân bố chuẩn tắc

Phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ với $\mu = 0$, $\sigma^2 = 1$ gọi là *phân bố chuẩn tắc* $\mathbf{N}(0;1)$.

Hàm mật độ xác suất của phân bố $\mathbf{N}(0;1)$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \; ; \quad \forall x \in \mathbb{R} \tag{2.39}$$

Hàm phân bố xác suất của $\mathbf{N}(0;1)$

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \; ; \quad \forall x \in \mathbb{R} \tag{2.40}$$

Có bảng tính sẵn các giá trị của $\varphi(x)$ và $\Phi(x)$ (xem Phụ lục 1 và Phụ lục 2).

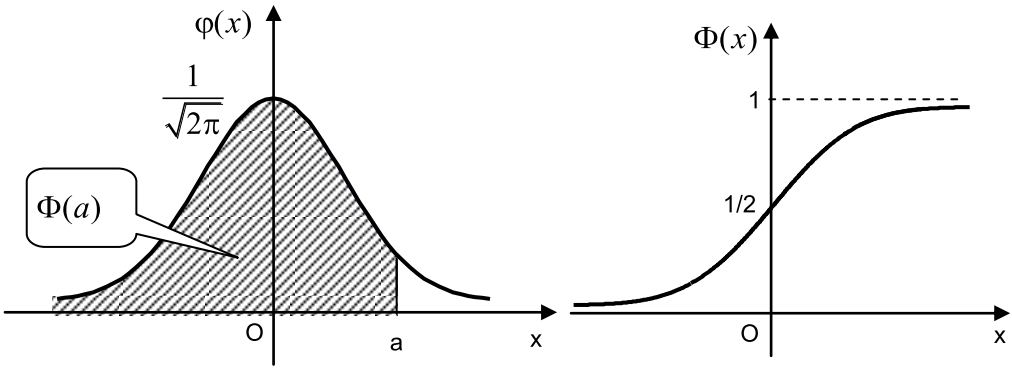
Cần chú ý rằng một số tài liệu cho bảng tính:

$$\Phi_0(x) = \int_{0}^{x} \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{0}^{x} e^{-\frac{t^2}{2}} dt \text{ với } x \geq 0 .$$

Công thức liên hệ là:

$$\Phi(x) = \Phi_0(x) + 0,5 \text{ với mọi } x \geq 0 .$$

Đồ thị của hàm mật độ xác suất $\varphi(x)$ và hàm phân bố xác suất $\Phi(x)$



a) Đồ thị hàm mật độ của phân bố chuẩn tắc $\mathbf{N}(0;1)$

b) Đồ thị hàm phân bố của phân bố chuẩn tắc $\mathbf{N}(0;1)$

Hình 2.13

Các tính chất của hàm phân bố xác suất $\Phi(x)$:

$$\Phi(x) + \Phi(-x) = 1, \quad \Phi(-x) = 1 - \Phi(x) .$$

Nếu $X \sim \mathbf{N}(0;1)$ thì:

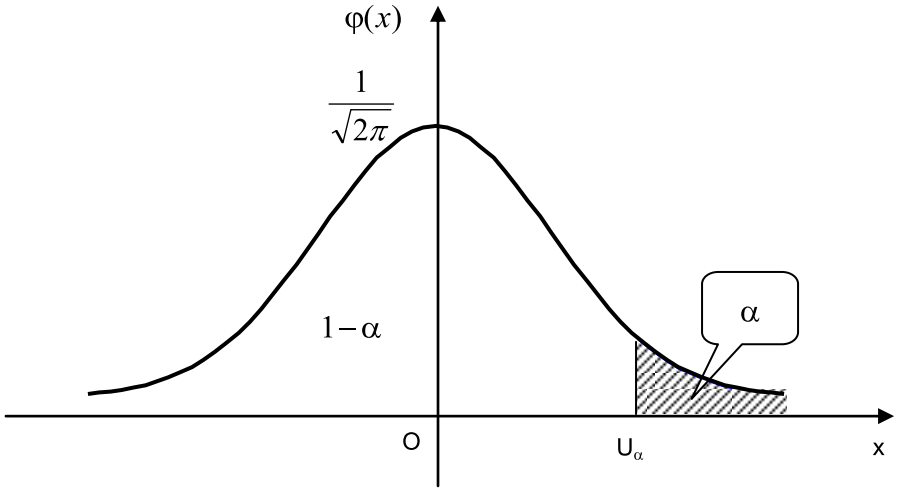
$$\forall a > 0,\ P\{|X| < a\} = 2\Phi(a) - 1,\ P\{|X| > a\} = 2\big(1 - \Phi(a)\big) . \quad (2.41)$$

Định nghĩa 2.12: *Giá trị* U_α *gọi là giá trị tới hạn mức* α *của phân bố chuẩn tắc nếu:*

$$\Phi^{-1}(1-\alpha) = U_\alpha . \quad (2.42)$$

Từ (2.41); (2.42) ta có: nếu $X \sim \mathbf{N}(0;1)$ thì:

$$P\{X > U_\alpha\} = \alpha ;\ P\left\{|X| > U_{\frac{\alpha}{2}}\right\} = \alpha ;\ P\left\{|X| < U_{\frac{\alpha}{2}}\right\} = 1-\alpha . \quad (2.43)$$



Hình 2.14 Giá trị tới hạn mức α *của phân bố chuẩn tắc* $\mathbf{N}(0;1)$

Ta có thể chứng minh được (xem ví dụ 3.17):

Nếu $X \sim \mathbf{N}(\mu;\sigma^2)$ thì $\dfrac{X-\mu}{\sigma} \sim \mathbf{N}(0;1)$. (2.44)

Từ đó ta có:

$$F_X(x) = P\{X \le x\} = P\left\{\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{2.45}$$

$$P\{a < X \le b\} = P\left\{\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right\}$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \tag{2.46}$$

Ví dụ 2.17: Giả sử $X \sim \mathbf{N}(\mu;\sigma^2)$, $\mu = 2100$, $\sigma = 200$. Hãy tìm:

a) $P\{X \le 2400\}$.

b) $P\{1700 < X \le 2200\}$.

c) Xác định a để $P\{X > a\} = 0,03$.

Giải: Áp dụng công thức (2.44), (2.45), (2.46) ta có:

a) $P\{X \le 2400\} = \Phi\left(\dfrac{2400-2100}{200}\right) = \Phi(1,5) = 0,9332$.

b) $P\{1700 < X \le 2200\} = \Phi\left(\dfrac{2200-2100}{200}\right) - \Phi\left(\dfrac{1700-2100}{200}\right)$

$= \Phi(0,5) - \Phi(-2) = 0,6688$.

c) $P\{X > a\} = 1 - \Phi\left(\dfrac{a-2100}{200}\right) = 0,03 \Rightarrow \Phi\left(\dfrac{a-2100}{200}\right) = 0,97$

Tra bảng ta được:

$$0,97 = \Phi(1,881) \Rightarrow \frac{a-2100}{200} = 1,881 \Rightarrow a = 2476,2.$$

2.3.2.4.4 Quy tắc hai xích ma và ba xích ma

Áp dụng công thức (2.41), (2.44) ta có thể tính xác suất của sự sai lệch giữa biến ngẫu nhiên có phân bố chuẩn $X \sim \mathbf{N}(\mu;\sigma^2)$ và tham số μ của nó theo công thức:

$$P\{|X-\mu| < \varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1 \qquad (2.47)$$

Nếu trong công thức (2.47) ta đặt $\varepsilon = 2\sigma$ tức là bằng hai lần độ lệch chuẩn của X (xem mục 2.47) thì:

$P\{|X-\mu| < 2\sigma\} = 2\Phi(2) - 1 = 0,9546$. Vậy:

$$P\{\mu - 2\sigma < X < \mu + 2\sigma\} = 0,9546 \qquad (2.48)$$

Tương tự thay $\varepsilon = 3\sigma$ ta được:

$$P\{\mu - 3\sigma < X < \mu + 3\sigma\} = 0,9974 \qquad (2.49)$$

Hai công thức trên là cơ sở của quy tắc hai xích ma và ba xích ma:

Nếu X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì có đến 95,46% giá trị của X nằm trong khoảng $(\mu - 2\sigma;\ \mu + 2\sigma)$ và hầu như toàn bộ giá trị của X nằm trong khoảng $(\mu - 3\sigma;\ \mu + 3\sigma)$.

2.3.2.5 Phân bố "khi bình phương"

Định nghĩa 2.13: *Biến ngẫu nhiên liên tục X có phân bố "khi bình phương" n bậc tự do, ký hiệu $X \sim \chi^2_n$ nếu hàm mật độ xác suất có dạng:*

$$f_X(x)=\begin{cases}\dfrac{x^{n/2-1}}{2^{n/2}\Gamma(n/2)}e^{-\frac{x}{2}} & \text{nếu } x>0\\ 0 & \text{nếu } x\le 0\end{cases} \tag{2.50}$$

trong đó $\Gamma(x)=\int_0^{+\infty} t^{x-1}e^{-t}dt$, $x>0$ là hàm Gamma.

Phân bố χ^2 do Karl Pearson đưa ra vào năm 1900.

Có thể chứng minh được rằng nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố chuẩn tắc $\mathbf{N}(0,1)$ thì:

$$\sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2_n \tag{2.51}$$

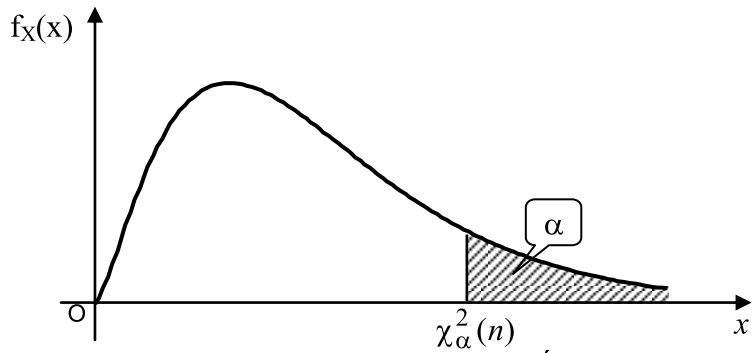
Từ (2.51) suy ra rằng nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân bố "khi bình phương" lần lượt n_1 và n_2 bậc tự do thì $X_1 + X_2$ là biến ngẫu nhiên có phân bố "khi bình phương" $n_1 + n_2$ bậc tự do

$$X_1 + X_2 \sim \chi^2_{n_1+n_2} \tag{2.52}$$

Giá trị tới hạn "khi bình phương" n bậc tự do mức α, ký hiệu $\chi^2_\alpha(n)$, được định nghĩa như sau:

$$P\left\{\chi^2 > \chi^2_\alpha(n)\right\} = \alpha\,. \tag{2.53}$$

Bảng các giá trị tới hạn $\chi^2_\alpha(n)$ được tính sẵn trong bảng ở Phụ lục 4.



Hình 2.15: Giá trị tới hạn của phân bố "khi bình phương"

2.3.2.6 Phân bố Student T(*n*)

Định nghĩa 2.14: *Biến ngẫu nhiên liên tục* T *có phân bố Student* n *bậc tự do, ký hiệu* $T \sim \mathbf{T}(n)$, *nếu hàm mật độ xác suất có dạng:*

$$f_T(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(n/2\right)}\left(1+\frac{x^2}{n}\right)^{\frac{-(n+1)}{2}}, \quad -\infty < x < \infty. \qquad (2.54)$$

trong đó $\Gamma(x)$ *là hàm Gamma.*

Người ta chứng minh được rằng nếu $Z \sim \mathbf{N}(0;1)$, $V \sim \chi^2_n$; Z và V độc lập thì:

$$T = \frac{Z}{\sqrt{V/n}} \sim \mathbf{T}(n). \qquad (2.55)$$

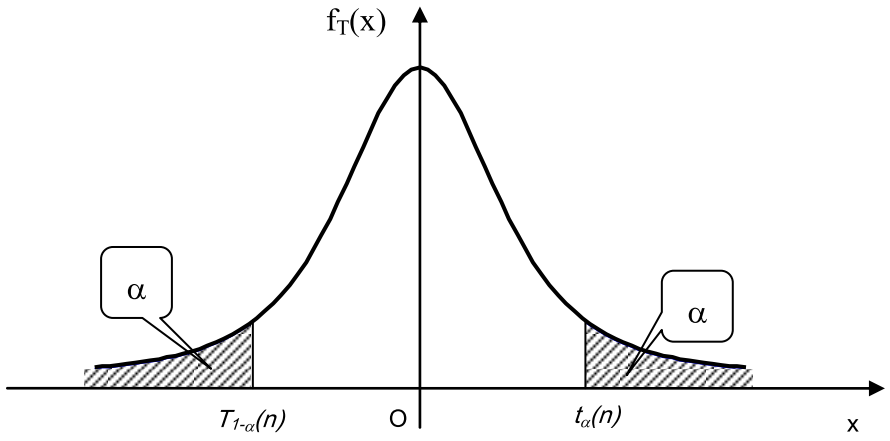
Giá trị tới hạn mức α của phân bố Student n bậc tự do ký hiệu $t_\alpha(n)$ thỏa mãn:

$$P\{T > t_\alpha(n)\} = \alpha. \qquad (2.56)$$

Bảng tính các giá trị tới hạn $t_\alpha(n)$ cho trong Phụ lục 3.

Hàm mật độ xác suất (2.54) là hàm chẵn nên đồ thị đối xứng qua trục tung. Khi số bậc tự do tăng lên, phân bố Student hội tụ rất nhanh

về phân bố chuẩn tắc $\mathbf{N}(0,1)$. Do đó khi n đủ lớn ($n \geq 30$) có thể dùng phân bố chuẩn tắc thay cho phân bố Student. Tuy nhiên khi n nhỏ ($n < 30$) việc thay thế như trên sẽ gặp sai số lớn.



Hình 2.16: Giá trị tới hạn của phân bố Student

2.4 CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

2.4.1 Kỳ vọng toán

2.4.1.1 Định nghĩa

Định nghĩa 2.15: *Với mọi biến ngẫu nhiên X ta ký hiệu $\mathrm{E}X$ hoặc $\mathrm{E}(X)$ xác định như sau:*

(i) Nếu X rời rạc nhận các giá trị x_i với hàm khối lượng xác suất $p_X(x_i)$ thì:

$$\mathrm{E}X = \sum_i x_i p_X(x_i) \tag{2.57}$$

(ii) Nếu X liên tục có hàm mật độ xác suất $f_X(x)$ thì:

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f_X(x)dx \tag{2.58}$$

Nếu chuỗi (2.57) hội tụ tuyệt đối (trường hợp X rời rạc) hoặc tích phân (2.58) hội tụ tuyệt đối (trường hợp X liên tục) thì ta gọi $\mathrm{E}X$ *là kỳ vọng của biến ngẫu nhiên* X (expected value), trường hợp ngược lại ta nói X không tồn tại kỳ vọng.

Kỳ vọng mang ý nghĩa là giá trị trung bình (average, mean value) của biến ngẫu nhiên X.

Ví dụ 2.18: Tính kỳ vọng của biến ngẫu nhiên X cho ở ví dụ 2.8.

Giải: $\mathrm{E}X = 0\times\frac{5}{30}+1\times\frac{15}{30}+2\times\frac{9}{30}+3\times\frac{1}{30}=\frac{6}{5}$.

Ví dụ 2.19: Theo thống kê việc một người Mỹ 25 tuổi sẽ sống thêm trên một năm có xác suất là 0,992, còn xác suất để người đó chết trong vòng một năm tới là 0,008 (xem ví dụ 1.17). Một chương trình bảo hiểm đề nghị người đó bảo hiểm sinh mạng cho 1 năm với số tiền chi trả 1000USD, còn tiền đóng là 10USD. Hỏi lợi nhuận của công ty bảo hiểm nhận được là bao nhiêu?

Giải: Rõ ràng lợi nhuận là biến ngẫu nhiên X với 2 giá trị là $+10$ đô la (nếu người bảo hiểm không chết) và $-990\,\mathrm{USD}$ (nếu người đó chết). Bảng phân bố xác suất tương ứng.

X	-990	+10
P	0,008	0,992

Do đó kỳ vọng $\mathrm{E}X = (-990)\cdot 0{,}008 + 10\cdot 0{,}992 = 2$. Ta thấy lợi nhuận trung bình là một số dương vì vậy công ty bảo hiểm có thể làm ăn có lãi.

Ví dụ 2.20: Tuổi thọ của một loại côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ xác suất như sau:

$$f_X(x)=\begin{cases} kx^2(4-x) & \text{nếu } 0\le x\le 4 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Tìm hàm phân bố xác suất và tìm tuổi thọ trung bình của loài côn trùng trên.

Giải: Vì $\int_0^4 x^2(4-x)dx = \frac{64}{3} \Rightarrow k = \frac{3}{64}$. Vậy hàm phân bố xác suất

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0 & \text{nếu} \quad x \le 0 \\ \frac{3x^3}{64}\left(\frac{4}{3} - \frac{x}{4}\right) & \text{nếu} \quad 0 < x \le 4 \\ 1 & \text{nếu} \quad x > 4 \end{cases}$$

Tuổi thọ trung bình $\mathrm{E}\,X = \frac{3}{64}\int_0^4 x^3(4-x)dx = \frac{3}{64}\left(x^4 - \frac{x^5}{5}\right)\Big|_0^4 = \frac{12}{5}$ (tháng).

2.4.1.2 Ý nghĩa của kỳ vọng

Kỳ vọng mang ý nghĩa là giá trị trung bình mà biến ngẫu nhiên nhận được. Giả sử biến ngẫu nhiên X nhận các giá trị $x_1, x_2, ..., x_m$ với các tần số tương ứng $r_1, r_2, ..., r_m$.

$r_i x_i$ là tổng giá trị X nhận được với cùng giá trị x_i.

Do đó $r_1x_1 + r_2x_2 + \cdots + r_mx_m$ là tổng tất cả các giá trị X nhận được.

$\frac{r_1x_1 + r_2x_2 + \cdots + r_mx_m}{n}$ là giá trị trung bình của X, trong đó $r_1 + r_2 + \cdots + r_m = n$.

Đặt $f_i = \frac{r_i}{n}$ là tần suất nhận giá trị x_i của X.

$$\frac{r_1x_1 + r_2x_2 + \cdots + r_mx_m}{n} = f_1x_1 + f_2x_2 + \cdots + f_mx_m.$$

Trong trường hợp tổng quát thì tần suất f_i được thay bằng xác suất p_i.

Trường hợp biến ngẫu nhiên liên tục phép tính tổng của giá trị trung bình được thay bằng phép tính tích phân xác định.

Khái niệm kỳ vọng được áp dụng rộng rãi trong nhiều lĩnh vực. Trong kinh doanh và quản lý, kỳ vọng được ứng dụng dưới dạng lợi nhuận kỳ vọng hay doanh số kỳ vọng.

2.4.1.3 Tính chất

1) $\mathrm{E}(C) = C$ với mọi hằng số C. (2.59)

2) $\mathrm{E}(CX) = C\mathrm{E}(X)$ với mọi hằng số C. (2.60)

3) $\mathrm{E}(X_1 + \cdots + X_n) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n)$. (2.61)

4) Cho hàm số $g(x)$, kỳ vọng của biến ngẫu nhiên $Y = g(X)$ được tính theo công thức:

$$\mathrm{E}\,Y = \begin{cases} \sum\limits_i g(x_i)p_X(x_i) & \text{nếu } X \text{ rời rạc với } p_X(x_i) = P\{X = x_i\} \\ \int\limits_{-\infty}^{\infty} g(x)f_X(x)dx & \text{nếu } X \text{ liên tục có hàm mật độ } f_X(x) \end{cases} \quad (2.62)$$

Đặc biệt ta có công thức tính kỳ vọng của X^2:

$$\mathrm{E}\,X^2 = \begin{cases} \sum\limits_i x_i^2 p_X(x_i) & \text{nếu } X \text{ rời rạc} \\ \int\limits_{-\infty}^{\infty} x^2 f_X(x)dx & \text{nếu } X \text{ liên tục có hàm mật độ } f_X(x) \end{cases} \quad (2.63)$$

5) Nếu $X_1, ..., X_n$ độc lập thì:

$$\mathrm{E}(X_1 \cdots X_n) = \mathrm{E}(X_1) \cdots \mathrm{E}(X_n). \quad (2.64)$$

Các tính chất (2.61); (2.64) được chứng minh trong chương 3.

Ví dụ 2.21: Chọn ngẫu nhiên 3 bi từ một túi có 6 bi đen, 4 bi trắng. Xét hai bài toán sau:

a) Nếu chọn được 1 bi trắng sẽ được thưởng 200USD. Gọi Y là số tiền nhận được. Tính kỳ vọng của Y.

b) Nếu chọn được 1 bi trắng sẽ được thưởng 200USD và chọn được 1 bi đen sẽ được thưởng 300USD. Gọi Z là số tiền nhận được. Tính kỳ vọng của Z.

Giải: a) Gọi X là số bi trắng trong 3 bi vừa chọn (xem ví dụ 2.6) thì $Y = g(X) = 200X$ là một biến ngẫu nhiên rời rạc có bảng phân bố sau:

Y = g(X)	0	200	400	600
P	5/30	15/30	9/30	1/30

$$\mathrm{E}\,Y = 0\times\frac{5}{30} + 200\times\frac{15}{30} + 400\times\frac{9}{30} + 600\times\frac{1}{30} = 240\,.$$

Mặt khác, theo công thức (2.60) và ví dụ 2.15 ta cũng được $\mathrm{E}\,Y = 200\mathrm{E}\,X = 200\times\frac{6}{5} = 240\,.$

b) $Z = 200X + 300(3 - X) = 900 - 100X$

$$\Rightarrow\ \mathrm{E}\,Z = \mathrm{E}\left(900 - 100X\right) = 900 - 100\mathrm{E}X = 900 - 100\times\frac{6}{5} = 780\$\,.$$

Ví dụ 2.22: Tung con xúc xắc n lần. Tìm kỳ vọng của tổng số chấm thu được.

Giải: Gọi $X_i\ (i = 1,...,n)$ là số chấm thu được ở lần tung thứ i, gọi X là tổng số chấm thu được trong n lần tung. Như vậy $X = \sum_{i=1}^{n} X_i$. Theo công thức (2.61) ta có $\mathrm{E}\,X = \sum_{i=1}^{n} \mathrm{E}\,X_i$.

Các biến ngẫu nhiên X_i đều có bảng phân bố xác suất như sau:

X_i	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

Do đó: $\mathrm{E}\,X_i = \frac{1}{6}\left(1+2+3+4+5+6\right) = \frac{7}{2}\ \Rightarrow\ \mathrm{E}\,X = \frac{7}{2}n\,.$

Ví dụ 2.23: (Bài toán thi đố). Ví dụ này khi mở rộng một cách thích hợp sẽ là mô hình về sự lựa chọn thứ tự tối ưu thỏa mãn điều kiện cho trước.

Xét trò chơi trả lời 2 câu hỏi A và B; người chơi có quyền chọn câu hỏi nào để trả lời đầu tiên. Câu hỏi A được trả lời đúng với xác suất 0,8 và khi đó người chơi sẽ được thưởng 100USD, câu hỏi B được trả lời đúng với xác suất 0,5 và người chơi được thưởng 200USD. Nếu không trả lời đúng lần thứ nhất sẽ không được trả lời tiếp. Vậy người chơi nên chọn câu hỏi nào trả lời đầu tiên để tiền thưởng trung bình nhận được cao hơn.

Gọi X là số tiền thưởng nhận được khi người chơi chọn câu hỏi A trả lời đầu tiên thì:

X	0USD	100 USD	300 USD
P	0,2	0,8.0,5	0,8.0,5

Do đó: $\mathrm{E}X = 0,4.100 + 0,4.300 = 160\mathrm{USD}$

Gọi Y là số tiền thưởng nhận được khi người chơi chọn câu hỏi B trả lời đầu tiên thì:

Y	0 USD	200 USD	300 USD
P	0,5	0,5.0,2	0,5.0,8

Do đó: $\mathrm{E}Y = 0,1.200 + 0,4.300 = 140\mathrm{USD}$

Vậy nên chọn câu hỏi A để trả lời đầu tiên để có khả năng nhận thưởng cao hơn.

2.4.2 Phương sai

2.4.2.1 Định nghĩa

Định nghĩa 2.16: *Phương sai (variance) hay độ lệch (deviation) bình phương trung bình của biến ngẫu nhiên X là đại lượng đo sự phân tán bình phương trung bình của X xung quanh giá trị trung bình EX. Nói cách khác phương sai của X là kỳ vọng của $\left(X - \mathrm{E}X\right)^2$.*

Phương sai của X được ký hiệu $\mathrm{D}X$ hoặc $\mathrm{Var}\,X$.

Vậy: $\mathrm{D}X = \mathrm{E}\big(X - \mathrm{E}\,X\big)^2$ (2.65)

Độ lệch chuẩn của X là: $\sigma_X = \sqrt{\mathrm{D}X}$.

Khai triển vế phải công thức (2.65) và áp dụng các tính chất của kỳ vọng ta có thể tính phương sai theo công thức sau:

$$\mathrm{D}X = \mathrm{E}\,X^2 - \big(\mathrm{E}\,X\big)^2 \tag{2.66}$$

Sử dụng công thức (2.63), (2.66) ta suy ra công thức thường dùng để tính phương sai của biến ngẫu nhiên rời rạc và phương sai của biến ngẫu nhiên liên tục:

(i) Nếu X rời rạc nhận các giá trị x_i với hàm khối lượng xác suất $p_X(x_i)$ thì:

$$\mathrm{D}\,X = \sum_i \big(x_i - \mathrm{E}\,X\big)^2\, p_X(x_i) = \sum_i x_i^2 p_X(x_i) - \left(\sum_i x_i p_X(x_i)\right)^2 \tag{2.67}$$

(ii) Nếu X liên tục có hàm mật độ xác suất $f_X(x)$ thì:

$$\mathrm{D}\,X = \int_{-\infty}^{\infty} \big(x - \mathrm{E}\,X\big)^2 f_X(x)dx = \int_{-\infty}^{\infty} x^2 f_X(x)dx - \left(\int_{-\infty}^{\infty} x f_X(x)dx\right)^2 \tag{2.68}$$

Ví dụ 2.24: Tính phương sai của biến ngẫu nhiên xét trong ví dụ 2.19.

Giải: $\mathrm{E}\,X^2 = (-990)^2 \cdot 0,008 + 10^2 \cdot 0,992 = 7940$

$$\Rightarrow \mathrm{D}X = \mathrm{E}X^2 - \big(\mathrm{E}X\big)^2 = 7940 - 4 = 7936$$

$$\Rightarrow \sigma_X = \sqrt{\mathrm{D}X} = \sqrt{7936} \approx 89,08.$$

Điều này nói lên rằng mặc dù kinh doanh bảo hiểm có lãi nhưng rủi ro khá lớn.

Ví dụ 2.25: Tính phương sai và độ lệch chuẩn của biến ngẫu nhiên xét trong ví dụ 2.20.

Giải: $\mathrm{E}X^2 = \dfrac{3}{64}\int_0^4 x^4(4-x)dx = \dfrac{3}{64}\left(\dfrac{4x^5}{5} - \dfrac{x^6}{6}\right)\Big|_0^4 = \dfrac{32}{5}$

$$\Rightarrow \mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \frac{32}{5} - \left(\frac{12}{5}\right)^2 = \frac{16}{25} \Rightarrow \sigma_X = \frac{4}{5}.$$

Phương sai của biến ngẫu nhiên X là độ lệch bình phương trung bình quanh giá trị trung bình $\mathrm{E}X$. Trong kỹ thuật phương sai đặc trưng cho mức độ phân tán của các chi tiết gia công hay sai số của thiết bị. Trong quản lý và kinh doanh thì phương sai đặc trưng cho mức độ rủi ro của các quyết định.

Ví dụ 2.19 cho thấy đầu tư bảo hiểm cho những người 25 tuổi là có lãi, nhưng ví dụ 2.24 cho thấy rủi ro của bảo hiểm rất lớn.

2.4.2.2 Tính chất

1) $\mathrm{D}(a) = 0$ với mọi hằng số a. (2.69)

2) $\mathrm{D}(aX+b) = a^2\mathrm{D}(X)$ với mọi hằng số a, b. (2.70)

3) Nếu $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ độc lập và có các phương sai hữu hạn thì:

$$\mathrm{D}\left(a_1X_1 + \cdots + a_nX_n\right) = a_1^2\mathrm{D}\left(X_1\right) + \cdots + a_n^2\mathrm{D}\left(X_n\right). \quad (2.71)$$

Nói riêng: Nếu X, Y độc lập và $\mathrm{D}X$, $\mathrm{D}Y$ hữu hạn thì:

$$\mathrm{D}\left(X \pm Y\right) = \mathrm{D}X + \mathrm{D}Y.$$

Ví dụ 2.26: Tung con xúc sắc n lần độc lập nhau. Tìm phương sai của tổng số chấm xuất hiện.

Giải: Xét $\boldsymbol{X} = \sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{X_i}$ ở ví dụ 2.22. Vì các $X_i\ (i = 1,...,n)$ độc lập nhau, do đó theo công thức (2.71) ta có $\mathrm{D}X = \sum_{i=1}^{n} \mathrm{D}X_i$.

Mặt khác: $\mathrm{E}\,X_i = \frac{7}{2}$; $\mathrm{E}\,X_i^2 = \frac{1}{6}\left(1^2+2^2+3^2+4^2+5^2+6^2\right) = \frac{91}{6}$

Do đó: $\mathrm{D}X_i = \frac{91}{6} - \frac{7^2}{2^2} = \frac{35}{12}$.

Vậy: $\mathrm{D}X = \frac{35}{12}n$.

2.4.3 Phân vị, Trung vị

2.4.3.1 Phân vị

Phân vị mức α của biến ngẫu nhiên X, ký hiệu v_α, là giá trị phân chia miền giá trị của X thỏa mãn

$$P\{X < v_\alpha\} \le \alpha \le P\{X \le v_\alpha\}$$

Nghĩa là: $F_X(v_\alpha^-) \le \alpha \le F_X(v_\alpha)$ (2.72)

- Trường hợp biến ngẫu nhiên X liên tục thì phân vị v_α là điểm phân chia miền giá trị của X thành 2 miền với xác suất tương ứng là α và $1-\alpha$. Vậy v_α là nghiệm duy nhất của phương trình $F_X(x) = \alpha$.

$$v_\alpha = F_X^{-1}(\alpha) \tag{2.73}$$

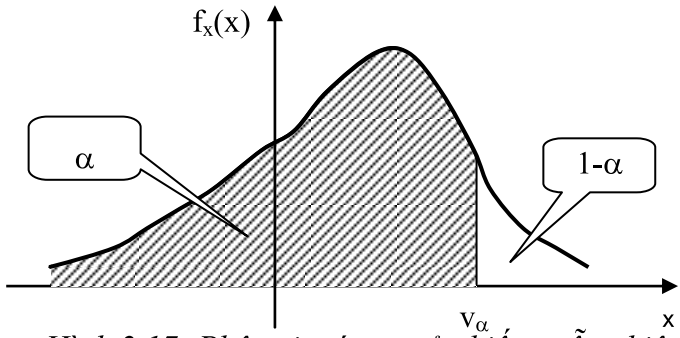
Giá trị tới hạn mức α là phân vị mức $1-\alpha$. (2.74)

- Trường hợp biến ngẫu nhiên X rời rạc có bảng phân bố:

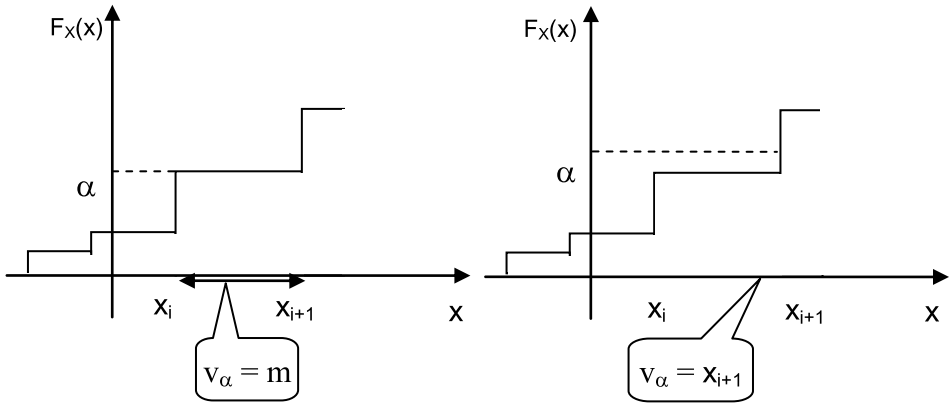
X	x_1	x_2	…
P	$p_X(x_1)$	$p_X(x_1)$	…

Đặt $P_i = p_X(x_1) + \cdots + p_X(x_i)$ thì:

$$v_\alpha = \begin{cases} m, \ \forall m \in [x_i, x_{i+1}] & \text{nếu } P_i = \alpha < P_{i+1} \\ x_{i+1} & \text{nếu } P_i < \alpha < P_{i+1} \end{cases} \tag{2.75}$$



Hình 2.17: Phân vị mức α của biến ngẫu nhiên liên tục



Hình 2.18: Phân vị mức α của biến ngẫu nhiên rời rạc

2.4.3.2 Trung vị

Phân vị mức 1/2 được gọi là *median* hay *trung vị* của X, ký hiệu $\text{Med}\,X$. Như vậy trung vị là điểm phân chia phân bố xác suất thành hai phần bằng nhau.

2.4.4 Mốt

Mốt (Mode) của biến ngẫu nhiên X là giá trị mà biến ngẫu nhiên X nhận với xác suất lớn nhất. Một biến ngẫu nhiên có thể có nhiều Mốt.

- Mốt của biến ngẫu nhiên rời rạc X với bảng phân bố:

X	x_1	x_2	...
P	$p_X(x_1)$	$p_X(x_2)$	...

được xác định như sau:

$$x_{i_0} = \text{Mod}\, X \;\Leftrightarrow\; p_X(x_{i_0}) = \max\{p_X(x_1), p_X(x_2), ...\} \quad (2.76)$$

- Nếu X liên tục có hàm mật độ xác suất $f(x)$ thì:

$$c = \text{Mod}\, X \;\Leftrightarrow\; f_X(c) = \max\{f_X(x),\ x \in \mathbb{3}\}. \quad (2.77)$$

Ví dụ 2.27: Biến ngẫu nhiên X ở ví dụ 2.8 có Mốt và trung vị $\text{Mod}\, X = \text{Med}\, X = 1$.

Ví dụ 2.28: Tìm trung vị và Mốt của biến ngẫu nhiên rời rạc có bảng phân bố xác suất:

X	20	21	22	23	24
P	0,3	0,25	0,18	0,14	0,13

Giải: Dễ thấy rằng $\text{Mod}\, X = 20$.

Hàm phân bố xác suất của X

$$F_X(x) = \begin{cases} 0 & \text{nếu} \quad x \le 20 \\ 0,3 & \text{nếu} \quad 20 < x \le 21 \\ 0,55 & \text{nếu} \quad 21 < x \le 22 \\ 0,73 & \text{nếu} \quad 22 < x \le 23 \\ 0,87 & \text{nếu} \quad 23 < x \le 24 \\ 1 & \text{nếu} \quad x > 24 \end{cases}$$

Từ đó suy ra $\text{Med}\, X = 21$; các giá trị $x \in [22;23]$ là phân vị mức 0,73 của X.

Ví dụ 2.29: Tìm $\text{Med}\, X$ và $\text{Mod}\, X$ của biến ngẫu nhiên liên tục X xét trong ví dụ 2.4

Giải: $\text{Med}\, X$ là nghiệm của phương trình:

$$F_X(x) = x^2 = \frac{1}{2} \;\Rightarrow\; \text{Med}\, X = \frac{1}{\sqrt{2}}.$$

Hàm mật độ xác suất $f_X(x)=\begin{cases}0 & \text{với} \quad x\le 0\\ 2x & \text{với} \quad 0<x\le 1\\ 0 & \text{với} \quad x>1\end{cases}$ đạt cực đại tại $x=1$, vậy $\text{Mod}\,X=1$.

Ví dụ 2.30: Tìm $\text{Med}\,X$ và $\text{Mod}\,X$ của biến ngẫu nhiên liên tục X có hàm mật độ xác suất xác định như sau

$$f_X(x)=\begin{cases}\dfrac{3}{4}x(2-x) & \text{với} \quad 0\le x\le 2\\ 0 & \text{nếu trái lại}\end{cases}$$

Giải: Hàm phân bố xác suất

$$F_X(x)=\begin{cases}0 & \text{với} \quad x\le 0\\ \dfrac{3}{4}\left(x^2-\dfrac{x^3}{3}\right) & \text{với} \quad 0<x\le 2\\ 1 & \text{với} \quad x>2\end{cases}$$

$\text{Med}\,X$ là nghiệm của phương trình:

$$F_X(x)=\frac{1}{2} \Leftrightarrow \begin{cases}x^3-3x^2+2=0\\ 0<x\le 2\end{cases}. \text{ Từ đó } \text{Med}\,X=1.$$

Hàm mật độ xác suất $f_X(x)$ có đạo hàm:

$$f_X'(x)=\begin{cases}\dfrac{3}{2}(1-x) & \text{với} \quad 0<x<2\\ 0 & \text{nếu trái lại}\end{cases}$$ đổi dấu từ dương sang âm khi đi qua $x=1$, do đó đạt cực đại tại điểm này. Vậy $\text{Mod}\,X=1$.

- Mốt của phân bố nhị thức (xem công thức (1.20))

$$X \sim \mathcal{B}(n;p) \text{ thì } \text{mod}\,X=\left[(n+1)p\right] \tag{2.78}$$

- Mốt của phân bố Poisson tham số $\lambda>0$

$$X \sim \mathcal{P}(\lambda) \text{ thì } \text{mod}\,X=\left[\lambda\right]. \tag{2.79}$$

2.4.5 Mô-men, hệ số bất đối xứng, hệ số nhọn (*)

1) Mô-men cấp k: $m_k = \mathrm{E}X^k$; $k = 1,2,...$ (2.80)

2) Mô-men quy tâm cấp k: $\mu_k = \mathrm{E}\left(X - \mathrm{E}X\right)^k$; $k = 1,2,...$ (2.81)

3) Hệ số bất đối xứng: $\alpha_3 = \dfrac{\mu_3}{\sigma^3}$ với $\sigma = \sqrt{\mathrm{D}X}$. (2.82)

4) Hệ số nhọn: $\alpha_4 = \dfrac{\mu_4}{\sigma^4}$. (2.83)

Nhận xét 2.2:

- $m_1 = \mathrm{E}X$, $\mu_1 = 0$, $\mu_2 = \mathrm{D}X$.

- α_3 đo mức độ bất đối xứng của luật phân bố:

Nếu $\alpha_3 < 0$ thì phân bố xác suất và đồ thị của hàm mật độ xác suất sẽ lệch về bên trái hơn.

$\alpha_3 = 0$ thì phân bố xác suất và đồ thị của hàm mật độ xác suất đối xứng.

$\alpha_3 > 0$ thì phân bố xác suất và đồ thị của hàm mật độ xác suất sẽ lệch về bên phải hơn.

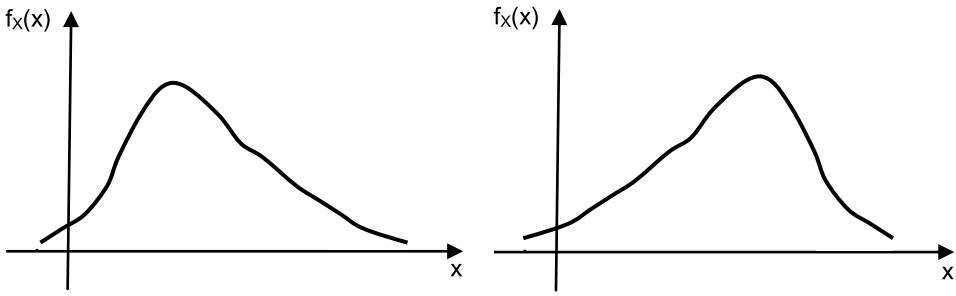
- Hệ số nhọn α_4 đặc trưng cho độ nhọn của đồ thị hàm mật độ xác suất so với đồ thị hàm mật độ xác suất của phân bố chuẩn.

Với biến ngẫu nhiên có phân bố chuẩn thì $\alpha_4 = 3$.

$\alpha_4 > 3$ thì đồ thị hàm mật độ xác suất sẽ nhọn hơn so với đồ thị hàm mật độ xác suất chuẩn.

$\alpha_4 < 3$ thì đồ thị hàm mật độ xác suất sẽ tù hơn so với đồ thị hàm mật độ xác suất chuẩn.

- Khi phân bố của X đối xứng hoặc gần đối xứng thì dùng kỳ vọng để định vị là tốt nhất, song nếu phân bố của X quá lệch thì nên dùng Trung vị và Mốt để định vị.



a) Hệ số bất đối xứng $\alpha_3 < 0$ *b) Hệ số bất đối xứng* $\alpha_3 > 0$

Hình 2.19

2.5 HÀM ĐẶC TRƯNG

2.5.1 Định nghĩa

Định nghĩa 2.17: *Hàm đặc trưng biến ngẫu nhiên* X *ký hiệu là* $\varphi_X(t)$ *và xác định như sau:*

$$\varphi_X(t) = \mathrm{E}\left(e^{itX}\right);\ i \text{ là đơn vị ảo thỏa mãn } i^2 = -1. \qquad (2.84)$$

- Hàm đặc trưng của biến ngẫu nhiên X rời rạc có bảng phân bố xác suất:

X	x_1	x_2	...
P	$p_X(x_1)$	$p_X(x_2)$	...

$$\varphi_X(t) = \mathrm{E}\left(e^{itX}\right) = \sum_k e^{itx_k} p_X(x_k). \qquad (2.85)$$

- Nếu X liên tục có hàm mật độ xác suất $f_X(x)$ thì:

$$\varphi_X(t) = \mathrm{E}\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx \qquad (2.86)$$

2.5.2 Tính chất

1) Hàm đặc trưng là biến đổi Fourier ngược của hàm mật độ xác suất:

$$\varphi_X(t) = \mathcal{F}^{-1}\{f_X(x)\}$$

Vì vậy hàm đặc trưng có các tính chất như biến đổi Fourier.

2) $|\varphi_X(t)| \le \varphi_X(0) = 1$

3) $\varphi_X(t)$ xác định không âm

4) Nếu biến ngẫu nhiên X có mô-men cấp k thì $\varphi_X(t)$ có đạo hàm đến cấp k và

$$\mathrm{E}\left[X^k\right] = \frac{1}{i^k}\varphi_X^{(k)}(0) \tag{2.87}$$

5) Các biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập khi và chỉ khi:

$$\varphi_{X_1+X_2+\cdots+X_n}(t) = \varphi_{X_1}(t)\cdots\varphi_{X_n}(t) \tag{2.88}$$

Nếu hàm đặc trưng $\varphi_X(t)$ có biến đổi Fourier (khả tích tuyệt đối và thỏa mãn điều kiện Dirichlet) thì hàm mật độ xác suất được tính theo công thức:

$$f_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi_X(t)dt \tag{2.89}$$

Ví dụ 2.31: Biến ngẫu nhiên X với phân bố Poisson tham số $\lambda > 0$ có hàm đặc trưng:

$$\varphi_X(t) = \sum_k e^{itx_k}p_k = \sum_{k=0}^{\infty} e^{itk}\frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\left(\lambda e^{it}\right)^k}{k!} = e^{\lambda(e^{it}-1)}$$

Ví dụ 2.32: Biến ngẫu nhiên X với phân bố mũ tham số $\lambda > 0$ có hàm đặc trưng:

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx}f(x)dx = \int_0^{\infty} e^{itx}\lambda e^{-\lambda x}dx$$

$$= \lambda\int_0^{\infty} e^{(it-\lambda)x}dx = \lambda\frac{e^{(it-\lambda)x}}{it-\lambda}\Bigg|_0^{\infty} = \frac{\lambda}{\lambda - it}.$$

2.5.3 Các đặc trưng của các quy luật phân bố xác suất thường gặp

Phân bố xác suất của X	Kỳ vọng	Phương sai	Hàm đặc trưng
Bernoulli tham số p	$\mathrm{E}\,X = p$	$\mathrm{D}\,X = pq$	$pe^{it} + q$
Nhị thức $\mathcal{B}(n;p)$	$\mathrm{E}X = np$	$\mathrm{D}X = npq$	$\left(pe^{it} + q\right)^n$
Poisson $\mathcal{P}(\lambda)$	$\mathrm{E}X = \lambda$	$\mathrm{D}X = \lambda$	$e^{\lambda(e^{it}-1)}$
Đều $\mathbf{U}(a,b)$	$\mathrm{E}X = \dfrac{a+b}{2}$	$\mathrm{D}X = \dfrac{(b-a)^2}{12}$	$\dfrac{e^{ibt} - e^{iat}}{it(b-a)}$
Phân bố mũ tham số $\lambda > 0$	$\mathrm{E}X = \dfrac{1}{\lambda}$	$\mathrm{D}X = \dfrac{1}{\lambda^2}$	$\dfrac{\lambda}{\lambda - it}$
Phân bố Erlang tham số $(k;\lambda)$	$\mathrm{E}X = \dfrac{k}{\lambda}$	$\mathrm{D}X = \dfrac{k}{\lambda^2}$	$\dfrac{\lambda^k}{(\lambda - it)^k}$
Phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$	$\mathrm{E}X = \mu$	$\mathrm{D}X = \sigma^2$	$e^{i\mu t - \frac{1}{2}\sigma^2 t^2}$

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

2.1 Biến ngẫu nhiên luôn luôn nhận giá trị dương.

Đúng ☐ Sai ☐ .

2.2 Biến ngẫu nhiên rời rạc chỉ nhận một số hữu hạn các giá trị.

Đúng ☐ Sai ☐ .

2.3 Nếu biến ngẫu nhiên X rời rạc chỉ nhận các giá trị $x_1,...,x_n$ thì hệ các biến cố $\{X = x_1\}$, …, $\{X = x_n\}$ lập thành một hệ đầy đủ.

Đúng ☐ Sai ☐ .

2.4 Kỳ vọng của biến ngẫu nhiên rời rạc là giá trị nó lấy thường xuyên nhất.

Đúng ☐ Sai ☐ .

2.5 Kỳ vọng của tổng hai biến ngẫu nhiên luôn luôn bằng tổng các kỳ vọng của nó.

Đúng ☐ Sai ☐ .

2.6 Hai biến ngẫu nhiên có cùng kỳ vọng sẽ có cùng phương sai.

Đúng ☐ Sai ☐ .

2.7 Phương sai của tổng hai biến ngẫu nhiên rời rạc luôn luôn bằng tổng phương sai của nó.

Đúng ☐ Sai ☐ .

2.8 Biến ngẫu nhiên tồn tại phương sai thì cũng tồn tại kỳ vọng.

Đúng ☐ Sai ☐ .

2.9 Hàm mật độ xác suất $f(x)$ của biến ngẫu nhiên liên tục có tính chất $f(x) \geq 0$.

Đúng ☐ Sai ☐ .

2.10 Tổng của hai biến ngẫu nhiên phân bố theo quy luật nhị thức bất kỳ luôn luôn là một biến ngẫu nhiên phân bố theo quy luật nhị thức.

Đúng ☐ Sai ☐.

2.11 Biến ngẫu nhiên phân bố theo quy luật Poisson là biến ngẫu nhiên rời rạc nên chỉ nhận một số hữu hạn các giá trị.

Đúng ☐ Sai ☐.

2.12 Nếu X là biến ngẫu nhiên có phân bố Poisson tham số $\lambda > 0$ thì kỳ vọng, phương sai và Mốt của X đều bằng λ.

Đúng ☐ Sai ☐.

2.13 Nếu biến ngẫu nhiên X phân bố theo quy luật chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì xác suất sai lệch giữa X và kỳ vọng của nó thỏa mãn $P\{|X-\mu|<\varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1$.

Đúng ☐ Sai ☐.

2.14 Nếu biến ngẫu nhiên X phân bố theo quy luật chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì $\frac{X-\mu}{\sigma}$ phân bố theo quy luật chuẩn tắc $\mathbf{N}(0;1)$.

Đúng ☐ Sai ☐.

2.15 Biến ngẫu nhiên phân bố theo quy luật Student chỉ nhận những giá trị dương.

Đúng ☐ Sai ☐.

2.16 Xác định các hằng số a, b sao cho hàm số sau là hàm phân bố của biến ngẫu nhiên X nào đó

$$F(x) = \begin{cases} 1 - ae^{-x/b} & \text{nếu } x \geq 0 \\ 0 & \text{nếu } x < 0 \end{cases}$$

2.17 Biến ngẫu nhiên X có bảng phân bố

X	-5	2	3	4
P	0,4	0,3	0,1	0,2

Tính kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$.

2.18 Biến ngẫu nhiên rời rạc X nhận ba giá trị có thể có là x_1, x_2, x_3. Biết $x_1 = 0,6, x_2 = 4$ với xác suất tương ứng $p_1 = 0,3$, $p_2 = 0,5$ và có kỳ vọng $\mathrm{E}X = 8$. Tìm x_3 và p_3.

2.19 Cho X_1 và X_2 là hai biến ngẫu nhiên độc lập có bảng phân bố xác suất như sau:

X_1	2	3	5
P	0,3	0,5	0,2

X_2	1	4
P	0,2	0,8

a) Tính $\mathrm{E}X_1$; $\mathrm{E}X_2$; $\mathrm{D}X_1$; $\mathrm{D}X_2$.

b) Tính $\mathrm{E}(X_1 + X_2)$ và $\mathrm{D}(X_1 + X_2)$.

2.20 Cho X_1, X_2, X_3 là ba biến ngẫu nhiên độc lập có bảng phân bố xác suất như sau:

X_1	0	2
P	0,6	0,4

X_2	1	2
P	0,4	0,6

X_3	0	2
P	0,8	0,2

Lập $\overline{X} = \dfrac{X_1 + X_2 + X_3}{3}$. Tính $\mathrm{E}(\overline{X})$; $\mathrm{D}(\overline{X})$.

2.21 Hai biến ngẫu nhiên X, Y độc lập. Tính $\mathrm{D}(Z)$ với:

a) $Z = 2X + 3Y$. b) $Z = -3X + Y$.

Cho biết $\mathrm{D}(X) = 4$, $\mathrm{D}(Y) = 5$.

2.22 Biến ngẫu nhiên rời rạc X nhận các giá trị có thể có là $x_1 = -1; x_2 = 0; x_3 = 1$. Tìm các xác suất tương ứng $p_1; p_2; p_3$ biết rằng $\mathrm{E}(X) = 0,1$ và $\mathrm{D}(X) = 0,89$.

2.23 Xếp ngẫu nhiên 5 hành khách lên 3 toa tàu I, II, III. Gọi X là số khách lên toa I và Y là số khách lên toa II và III.

a) Tính xác suất để cả 3 toa đều có khách.

b) Lập bảng phân bố xác suất của biến ngẫu nhiên X và biến ngẫu nhiên Y.

2.24 Tính kỳ vọng của biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x)=\begin{cases}\dfrac{\cos x}{2} & \text{nếu } x\in(-\pi/2;\pi/2)\\ 0 & \text{nếu } x\notin(-\pi/2;\pi/2)\end{cases}$$

2.25 Tuổi thọ của một loài côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ xác suất như sau:

$$f_X(x)=\begin{cases}kx^2(2-x) & \text{nếu } 0\le x\le 2\\ 0 & \text{nếu trái lại}\end{cases}$$

a) Tìm k; Tính xác suất để côn trùng chết trước khi nó được một tháng tuổi;

b) Tìm $\mathrm{E}X$, $\mathrm{D}X$.

2.26 Hai xạ thủ $\mathscr{A}$ và $\mathscr{B}$ tập bắn. Mỗi người bắn hai phát. Xác suất bắn trúng đích của $\mathscr{A}$ trong mỗi lần bắn là 0,4; còn của $\mathscr{B}$ là 0,5.

a) Gọi X là số phát bắn trúng của $\mathscr{A}$ trừ đi số phát bắn trúng của $\mathscr{B}$. Tìm phân bố xác suất của X, kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$.

b) Tìm phân bố xác suất của $Y=|X|$ và kỳ vọng $\mathrm{E}Y$.

2.27 Một xí nghiệp có hai ôtô vận tải hoạt động. Xác suất trong ngày làm việc các ôtô bị hỏng tương ứng bằng 0,1 và 0,2. Gọi X là số ôtô bị hỏng trong thời gian làm việc. Lập bảng phân bố xác suất, tính kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$ của X.

2.28 Cho biến ngẫu nhiên rời rạc X có bảng phân bố xác suất

X	1	2	3	4	5	6	7
P	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$

a) Xác định k.

b) Tính xác suất $P\{X\ge 5\}$ và $P\{X<3\}$.

c) Tính kỳ vọng $\mathrm{E}X$.

d) Tính phương sai $\mathrm{D}X$.

2.29 Có 5 sản phẩm trong đó có 4 chính phẩm và 1 phế phẩm. Người ta lấy ra lần lượt 2 sản phẩm (lấy không hoàn lại).

a) Gọi X là "số phế phẩm có thể gặp phải". Lập bảng phân bố xác suất của X. Tính kỳ vọng EX và phương sai DX.

b) Gọi Y là "số chính phẩm có thể nhận được". Lập hệ thức cho biết mối quan hệ giữa Y và X. Tính kỳ vọng EY và phương sai DY.

2.30 Một nhóm có 10 người trong đó có 6 nam và 4 nữ. Chọn ngẫu nhiên ra 3 người. Gọi X là số nữ có trong nhóm được chọn. Lập bảng phân bố xác suất của X. Tính kỳ vọng EX.

2.31 Hai kiện tướng bóng bàn ngang sức thi đấu với nhau. Hỏi thắng 2 trong 4 ván dễ hơn hay thắng 3 trong 6 ván dễ hơn.

2.32 Một nữ công nhân quản lý 12 máy dệt. Xác suất để mỗi máy trong khoảng thời gian T cần đến sự chăm sóc của nữ công nhân là 1/3. Tính xác suất:

a) Trong khoảng thời gian T có 4 máy cần đến sự chăm sóc của nữ công nhân.

b) Trong khoảng thời gian T có từ 3 đến 6 máy cần đến sự chăm sóc của nữ công nhân.

2.33 Trong một lô hàng có 800 sản phẩm loại 1 và 200 sản phẩm loại 2. Lấy ngẫu nhiên ra 5 sản phẩm theo phương thức có hoàn lại. Gọi X là số sản phẩm loại 1 lấy được.

a) X tuân theo quy luật phân bố gì? Viết biểu thức tổng quát của quy luật.

b) Tìm kỳ vọng và phương sai của X.

c) Tìm mốt của X và tính khả năng để xảy ra điều đó.

2.34 Xác suất để sản phẩm sản xuất ra bị hỏng bằng 0,1.

a) Tìm xác suất để trong 5 sản phẩm sản xuất ra có không quá 2 sản phẩm hỏng.

b) Tìm số sản phẩm hỏng trung bình trong 5 sản phẩm đó.

c) Tìm số sản phẩm hỏng có khả năng xảy ra nhiều nhất.

2.35 Một bài thi trắc nghiệm gồm có 10 câu hỏi, mỗi câu hỏi có 5 phương án trả lời, trong đó chỉ có một phương án đúng. Giả sử mỗi câu trả lời đúng được 4 điểm và câu trả lời sai bị trừ 2 điểm. Một học sinh kém làm bài bằng cách chọn hú họa một phương án cho mỗi câu hỏi. Tính xác suất để:

a) Anh ta được 4 điểm.

b) Anh ta bị điểm âm.

2.36 Một nguồn nhị phân sinh ngẫu nhiên hai chữ số 1, 0 với xác suất tương ứng 0,6 và 0,4.

a) Tính xác suất có hai chữ số 1 và ba chữ số 0 trong một dãy có năm chữ số.

b) Tính xác suất có ít nhất ba chữ số 1 trong một dãy có năm chữ số.

2.37 Tín hiệu thông tin được phát đi 5 lần độc lập nhau. Xác suất thu được tin của mỗi lần phát là 0,7. Tính xác suất:

a) Thu được tín hiệu đúng 2 lần.

b) Thu được tín hiệu nhiều nhất 1 lần.

c) Thu được tin.

2.38 Một cầu thủ nổi tiếng về đá phạt đền, xác suất đá vào gôn là 4/5. Có người cho rằng cứ "sút" 5 quả thì chắc chắn rằng có 4 quả vào lưới. Điều khẳng định đó có đúng không? Tìm xác suất để trong 5 lần sút có đúng 4 lần bóng vào lưới.

2.39 Ở một tổng đài bưu điện các cuộc điện thoại gọi đến xuất hiện một cách ngẫu nhiên, độc lập với nhau và trung bình có 2 cuộc gọi trong một phút. Tính xác suất để:

a) Có ít nhất một cuộc gọi trong khoảng thời gian 10 giây.

b) Trong khoảng thời gian 3 phút có nhiều nhất ba cuộc gọi.

c) Trong khoảng thời gian 3 phút liên tiếp mỗi phút có nhiều nhất một cuộc gọi.

2.40 Giả sử X là biến ngẫu nhiên rời rạc với phân bố cấp số nhân tham số p có hàm khối lượng xác suất $p_X(x)$ thỏa mãn công thức (2.16).

a) Nghiệm lại công thức $\sum_{k=1}^{\infty} p_X(k) = 1$. Tìm hàm phân bố $F_X(x)$.

b) Tính kỳ vọng $\mathrm{E}\boldsymbol{X}$, phương sai $\mathrm{D}X$.

2.41 Biến ngẫu nhiên X tuân theo quy luật chuẩn với kỳ vọng $\mu = 10$ và phương sai $\sigma^2 = 4$. Tính xác suất để X nhận giá trị trong khoảng (8; 12).

2.42 Biến ngẫu nhiên X tuân theo quy luật chuẩn với kỳ vọng $\mu = 10$. Xác suất để X nhận giá trị trong khoảng (10; 20) là 0,3. Tìm xác suất để X nhận giá trị trong khoảng (0; 10).

2.43 Trọng lượng sản phẩm X do một máy tự động sản xuất là một biến ngẫu nhiên tuân theo quy luật chuẩn với $\mu = 100$ gam và độ lệch chuẩn $\sigma = 1$ gam. Sản phẩm được coi là đạt tiêu chuẩn kỹ thuật nếu trọng lượng của nó đạt từ 98 đến 102gam.

a) Tìm tỷ lệ sản phẩm đạt tiêu chuẩn kỹ thuật của nhà máy.

b) Tìm tỷ lệ phế phẩm của nhà máy.

c) Giải thích bằng đồ thị kết quả tìm được ở phần a).

2.44 Chiều cao của nam giới khi trưởng thành ở một vùng dân cư là biến ngẫu nhiên phân bố chuẩn với kỳ vọng $\mu = 160$ cm và độ lệch chuẩn $\sigma = 6$ cm. Một thanh niên bị coi là lùn nếu có chiều cao nhỏ hơn 155 cm.

a) Tìm tỷ lệ thanh niên lùn ở vùng đó.

b) Tìm xác suất để lấy ngẫu nhiên 4 người thì có ít nhất 1 người không bị lùn.

2.45 Cho X_i $(i=\overline{1,n})$ là các biến ngẫu nhiên độc lập, cùng tuân theo quy luật chuẩn với

$$\mathrm{E}(X_1)=\mathrm{E}(X_2)=...=\mathrm{E}(X_n)=\mu;$$

$$\mathrm{D}(X_1)=\mathrm{D}(X_2)=...=\mathrm{D}(X_n)=\sigma^2$$

Lập công thức tính $P\left\{\left|\overline{X}-\mu\right|<\varepsilon\right\}$ biết rằng $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ và cũng tuân theo quy luật chuẩn, $\varepsilon>0$ tùy ý.

2.46 Biến ngẫu nhiên liên tục X được gọi là *biến ngẫu nhiên Rayleigh* tham số σ nếu hàm mật độ có dạng

$$f_X(x)=\begin{cases}\dfrac{x}{\sigma^2}e^{-x^2/(2\sigma^2)} & x>0\\ 0 & x<0\end{cases}$$

a) Xác định hàm phân bố $F_X(x)$

b) Tính kỳ vọng $\mathrm{E}X$, phương sai $\mathrm{D}X$.

Chương 3

VÉC TƠ NGẪU NHIÊN

GIỚI THIỆU

Véc tơ ngẫu nhiên hay biến ngẫu nhiên nhiều chiều là một bộ có thứ tự bao gồm nhiều biến ngẫu nhiên, mỗi biến ngẫu nhiên là một thành phần của véc tơ ngẫu nhiên. Số biến ngẫu nhiên thành phần gọi là chiều của véc tơ ngẫu nhiên.

Tương tự biến ngẫu nhiên, quy luật phân bố xác suất của véc tơ ngẫu nhiên được khảo sát thông qua hàm phân bố xác suất. Véc tơ ngẫu nhiên có các biến ngẫu nhiên thành phần rời rạc được gọi là véc tơ ngẫu nhiên rời rạc. Nếu tất cả các biến ngẫu nhiên thành phần liên tục thì véc tơ ngẫu nhiên tương ứng gọi là liên tục. Véc tơ ngẫu nhiên rời rạc được xác định bởi bảng phân bố xác suất đồng thời, còn véc tơ ngẫu nhiên liên tục được xác định bởi hàm mật độ xác suất đồng thời.

Quy luật phân bố của véc tơ ngẫu nhiên hai chiều rời rạc được xác định bởi bảng phân bố xác suất đồng thời, đó là bảng ghi các giá trị của hai biến ngẫu nhiên thành phần theo hàng, theo cột và xác suất tương ứng. Dựa vào công thức xác suất đầy đủ ta có thể tìm được bảng phân bố xác suất của hai biến ngẫu nhiên thành phần. Tương tự, từ hàm mật độ xác suất đồng thời ta có thể tìm hàm mật độ xác suất của biến ngẫu nhiên thành phần bằng cách lấy tích phân theo các biến thích hợp.

Ngoài đặc trưng kỳ vọng và phương sai của các biến ngẫu nhiên thành phần, các véc tơ ngẫu nhiên còn được đặc trưng bởi hiệp phương sai và hệ số tương quan. Hệ số tương quan đo mức độ phụ

thuộc tuyến tính của hai biến ngẫu nhiên thành phần, hệ số tương quan càng gần 1 thì mức độ phụ thuộc tuyến tính càng chặt. Hai biến ngẫu nhiên thành phần không tương quan thì hệ số tương quan bằng 0. Dựa vào đặc trưng này ta có thể xây dựng hàm hồi quy tương quan. Hai biến ngẫu nhiên độc lập thì không tương quan.

Tính độc lập của các biến ngẫu nhiên được nhận biết thông qua dấu hiệu của bảng phân bố xác suất, hàm phân bố xác suất hoặc hàm mật độ xác suất.

Đối với biến ngẫu nhiên không độc lập, bằng cách áp dụng công thức tính xác suất có điều kiện có thể suy ra quy luật phân bố xác suất có điều kiện của các biến ngẫu nhiên thành phần.

Từ các biến ngẫu nhiên $X_1, X_2, ..., X_n$ và hàm nhiều biến $\varphi(x_1, ..., x_n)$ ta có thể xây dựng các biến ngẫu nhiên mới $\varphi(X_1, ..., X_n)$ đó là hàm của các biến ngẫu nhiên đã cho. Ta có thể xác định bảng phân bố xác suất hay hàm mật độ xác suất của $\varphi(X_1, ..., X_n)$. Mỗi thống kê của mẫu ngẫu nhiên đều có dạng này (xem chương 5).

Để học tốt chương này bạn đọc cần nắm vững các tính chất cơ bản của xác suất, xác suất có điều kiện, hàm phân bố xác suất. Tích phân của hàm nhiều biến.

3.1 KHÁI NIỆM VÉC TƠ NGẪU NHIÊN

3.1.1 Khái niệm

Trong nội dung chương 2 ta đã xét các biến ngẫu nhiên mà giá trị chúng nhận được có thể biểu diễn bằng giá trị số, đó là các biến ngẫu nhiên một chiều. Tuy nhiên trong thực tế có thể gặp các đại lượng ngẫu nhiên mà giá trị nhận được là các bộ gồm hai, ba,…, n số. Những đại lượng này được gọi một cách tương ứng là biến ngẫu nhiên hai chiều, ba chiều,…, n chiều và được gọi chung là véc tơ ngẫu nhiên.

Định nghĩa 3.1: *Một véc tơ ngẫu nhiên n chiều là một bộ có thứ tự $(X_1, X_2, ..., X_n)$ với các thành phần $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên xác định trong cùng một phép thử.*

Ta ký hiệu véc tơ ngẫu nhiên hai chiều là (X,Y)*, trong đó* X *là biến ngẫu nhiên thành phần thứ nhất và* Y *là biến ngẫu nhiên thành phần thứ hai.*

Ví dụ 3.1: Một nhà máy sản xuất một loại sản phẩm. Nếu kích thước của sản phẩm được đo bằng chiều dài X và chiều rộng Y thì ta có biến ngẫu nhiên hai chiều, còn nếu xét thêm cả chiều cao Z nữa thì ta có biến ngẫu nhiên ba chiều. Nếu ta chỉ quan tâm đến trọng lượng và thể tích của sản phẩm ta cũng được biến ngẫu nhiên hai chiều.

Véc tơ ngẫu nhiên n chiều $(X_1, X_2, ..., X_n)$ là liên tục hay rời rạc nếu tất cả các biến ngẫu nhiên thành phần $X_1, X_2, ..., X_n$ là liên tục hay rời rạc.

3.1.2 Hàm phân bố xác suất

Định nghĩa 3.2: *Hàm* n *biến* $F_{X_1 \dots X_n}(x_1, x_2, ..., x_n)$ *xác định bởi*:

$$F_{X_1 \dots X_n}(x_1, x_2, ..., x_n) = P\{X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n\} \quad (3.1)$$

được gọi là hàm phân bố xác suất của véc tơ ngẫu nhiên $X = (X_1, X_2, ..., X_n)$ *hay được gọi là hàm phân bố xác suất đồng thời của các biến ngẫu nhiên* $X_1, X_2, ..., X_n$.

Hàm phân bố xác suất đồng thời có các tính chất:

1. $0 \le F_{X_1 \dots X_n}(x_1, ..., x_n) \le 1$. (3.2)

2. $\lim\limits_{x_k \to -\infty} F_{X_1 \dots X_n}(x_1, ..., x_n) = 0$, với k nào đó thuộc $\{1, ..., n\}$. (3.3)

3. $\lim\limits_{(x_1, ..., x_n) \to (\infty, ..., \infty)} F_{X_1 \dots X_n}(x_1, ..., x_n) = 1$. (3.4)

4. $F_{X_1 \dots X_n}(x_1, ..., x_n)$ không giảm theo từng biến. (3.5)

5. $\lim\limits_{x_1 \to \infty} F_{X_1 \dots X_n}(x_1, ..., x_n) = F_{X_2 \dots X_n}(x_2, ..., x_n)$.

Như vậy nếu lấy giới hạn của hàm phân bố xác suất đồng thời của n biến ngẫu nhiên $X_1, X_2, ..., X_n$ khi biến x_1 tiến đến vô cùng thì

được hàm phân bố xác suất đồng thời của $n-1$ biến ngẫu nhiên $X_2,...,X_n$.

Tương tự nếu lấy giới hạn của hàm phân bố xác suất đồng thời của $X_1,X_2,...,X_n$ khi biến x_k tiến đến vô cùng, với k nào đó thuộc $\{1,2,...,n\}$, thì được hàm phân bố xác suất đồng thời của $n-1$ biến ngẫu nhiên $X_1,...,X_{k-1},X_{k+1},...,X_n$.

6. Đặc biệt nếu $F_{XY}(x,y)$ là hàm phân bố xác suất của véc tơ ngẫu nhiên hai chiều (X,Y) thì:

$$\begin{aligned}&\lim_{y\to\infty} F_{XY}(x,y)=P\{X\le x\}=F_X(x);\\&\lim_{x\to\infty} F_{XY}(x,y)=P\{Y\le y\}=F_Y(y).\end{aligned}\tag{3.6}$$

$F_X(x)$, $F_Y(y)$ là các hàm phân bố xác suất của biến ngẫu nhiên X, Y và được gọi là các hàm phân bố xác suất thành phần của véc tơ ngẫu nhiên (X,Y), hay còn gọi là hàm phân bố xác suất biên của hàm phân bố xác suất đồng thời $F_{XY}(x,y)$.

7. $$\begin{aligned}&P\{x_1<X\le x_2,y_1<Y\le y_2\}\\&=F_{XY}(x_2,y_2)-F_{XY}(x_1,y_2)-F_{XY}(x_2,y_1)+F_{XY}(x_1,y_1).\end{aligned}\tag{3.7}$$

Ví dụ 3.2: Hàm số $F(x,y)=\begin{cases}1-e^{-(x+y)} & 0\le x<\infty,0\le y<\infty\\ 0 & \text{nếu ngược lại}\end{cases}$

Xét $x_2=y_2=2$ và $x_1=y_1=1$, ta có:

$$\begin{aligned}&F(2,2)-F(1,2)-F(2,1)+F(1,1)\\&=(1-e^{-4})-(1-e^{-3})-(1-e^{-3})+(1-e^{-2})\\&=-e^{-4}+2e^{-3}-e^{-2}=-(e^{-2}-e^{-1})^2<0.\end{aligned}$$

Vậy $F(x,y)$ không thể là hàm phân bố của véc tơ ngẫu nhiên vì không thỏa mãn (3.7).

Ví dụ 3.3: Cho hàm phân bố xác suất của véc tơ ngẫu nhiên (X,Y) xác định như sau:

$$F_{XY}(x,y)=\begin{cases}(1-e^{-\alpha x})(1-e^{-\beta y}) & x\ge 0, y\ge 0; \alpha,\beta>0\\ 0 & \text{nếu ngược lại}\end{cases}$$

a) Tìm hai hàm phân bố xác suất biên $F_X(x)$, $F_Y(y)$.

b) Tính $P\{X\le 1, Y\le 1\}$, $P\{X\le 1\}$, $P\{Y>1\}$ và $P\{X>x, Y>y\}$.

Giải:

a) $F_X(x)=F_{XY}(x,\infty)=\begin{cases}1-e^{-\alpha x} & x\ge 0\\ 0 & x<0\end{cases}$;

$$F_Y(y)=F_{XY}(\infty,y)=\begin{cases}1-e^{-\beta y} & y\ge 0\\ 0 & y<0\end{cases}.$$

b) $P\{X\le 1, Y\le 1\}=F_{XY}(1,1)=(1-e^{-\alpha})(1-e^{-\beta})$;

$P\{X\le 1\}=F_X(1)=1-e^{-\alpha}$;

$P\{Y>1\}=1-P\{Y\le 1\}=1-F_Y(1)=e^{-\beta}$;

Áp dụng định luật De Morgan ta có:

$$\overline{\{X>x\}\{Y>y\}}=\overline{\{X>x\}}\cup\overline{\{Y>y\}}=\{X\le x\}\cup\{Y\le y\}$$

Do đó:

$$P\left(\overline{\{X>x\}\{Y>y\}}\right)=P\left(\{X\le x\}\cup\{Y\le y\}\right)$$
$$=P\{X\le x\}+P\{Y\le y\}-P\{X\le x; Y\le y\}$$
$$=F_X(x)+F_Y(y)-F_{XY}(x,y)$$
$$=(1-e^{-\alpha x})+(1-e^{-\beta y})-(1-e^{-\alpha x})(1-e^{-\beta y})=1-e^{-\alpha x}e^{-\beta y}.$$

Vậy: $P\{X>x, Y>y\}=1-P\left(\overline{\{X>x\}\{Y>y\}}\right)=e^{-\alpha x}e^{-\beta y}$.

3.2 BẢNG PHÂN BỐ XÁC SUẤT CỦA VÉC TƠ NGẪU NHIÊN RỜI RẠC HAI CHIỀU

Tương tự trường hợp biến ngẫu nhiên rời rạc, quy luật phân bố xác suất của véc tơ ngẫu nhiên rời rạc 2 chiều có thể được xác định thông qua bảng phân bố xác suất đồng thời.

3.2.1 Bảng phân bố xác suất đồng thời

Bảng phân bố xác suất đồng thời của biến ngẫu nhiên rời rạc hai chiều (X,Y) là bảng liệt kê tất cả các giá trị của X theo cột, Y theo hàng và các xác suất tương ứng.

X \ Y	y_1	y_2	...	y_j	...	y_m	$\sum_j$
x_1	$p_{XY}(x_1,y_1)$	$p_{XY}(x_1,y_2)$	...	$p_{XY}(x_1,y_j)$	...	$p_{XY}(x_1,y_m)$	$p_X(x_1)$
x_2	$p_{XY}(x_2,y_1)$	$p_{XY}(x_2,y_2)$	...	$p_{XY}(x_2,y_j)$	...	$p_{XY}(x_2,y_m)$	$p_X(x_2)$
⋮	⋮	⋮	...	⋮	...	⋮	⋮
x_i	$p_{XY}(x_i,y_1)$	$p_{XY}(x_i,y_2)$	...	$p_{XY}(x_i,y_j)$	...	$p_{XY}(x_i,y_m)$	$p_X(x_i)$
⋮	⋮	⋮	...	⋮	...	⋮	⋮
x_n	$p_{XY}(x_n,y_1)$	$p_{XY}(x_n,y_2)$	...	$p_{XY}(x_n,y_j)$	...	$p_{XY}(x_n,y_m)$	$p_X(x_n)$
$\sum_i$	$p_Y(y_1)$	$p_Y(y_2)$	...	$p_Y(y_j)$	...	$p_Y(y_m)$	1

Trong đó x_i $(i=\overline{1,n})$ là các giá trị có thể có của thành phần X; y_j $(j=\overline{1,m})$ là các giá trị có thể có của thành phần Y. $p_{XY}(x_i,y_j)$ là *hàm khối lượng xác suất đồng thời* của biến ngẫu nhiên hai chiều (X,Y) xác định bởi:

$$p_{XY}(x_i,y_j)=P\left\{X=x_i,Y=y_j\right\}, \tag{3.8}$$

thỏa mãn điều kiện:

$$\begin{cases} p_{XY}(x_i,y_j)\geq 0,\ \forall i=\overline{1,n}, j=\overline{1,m} \\ \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} p_{XY}(x_i,y_j)=1 \end{cases} \tag{3.9}$$

Hàm phân bố xác suất đồng thời xác định như sau:

$$F_{XY}(x,y)=\sum_{x_i\leq x}\sum_{y_j\leq y} p_{XY}(x_i,y_j) \tag{3.10}$$

3.2.2 Bảng phân bố xác suất biên

Áp dụng công thức xác suất đầy đủ (1.19) cho hệ $\{X = x_1\}$, $\{X = x_2\}$, ..., $\{X = x_n\}$ (xem công thức 2.14) ta có:

$$p_Y(y_j) = P\{Y = y_j\} = \sum_{i=1}^{n} P\{X = x_i, Y = y_j\} = \sum_{i=1}^{n} p_{XY}(x_i, y_j); j = \overline{1,m} \quad (3.11)$$

$$p_X(x_i) = P\{X = x_i\} = \sum_{j=1}^{m} P\{X = x_i, Y = y_j\} = \sum_{j=1}^{m} p_{XY}(x_i, y_j); i = \overline{1,n} \quad (3.12)$$

Như vậy từ bảng phân bố xác suất đồng thời của (X,Y), nếu ta cộng các xác suất theo cột thì ta được các xác suất tương ứng với các giá trị của Y, nếu ta cộng các xác suất theo hàng ta được các xác suất tương ứng với giá trị của X. Từ đó nhận được phân bố xác suất của biến ngẫu nhiên thành phần Y và biến ngẫu nhiên thành phần X.

X	x_1	x_2	...	x_i	...	x_n
P	$p_X(x_1)$	$p_X(x_2)$	...	$p_X(x_i)$	...	$p_X(x_n)$

Y	y_1	y_2	...	y_j	...	y_m
P	$p_Y(y_1)$	$p_Y(y_2)$	...	$p_Y(y_j)$	...	$p_Y(y_m)$

Ví dụ 3.4: Gieo 3 đồng tiền cân đối $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$. Gọi X là số mặt ngửa xuất hiện của 2 đồng tiền $\mathscr{A}$, $\mathscr{B}$ và Y là số mặt ngửa xuất hiện của cả 3 đồng tiền $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$. Hãy lập bảng phân bố xác suất đồng thời của X,Y.

Giải: Chúng ta có bảng 8 kết quả đồng khả năng của việc gieo 3 đồng tiền cân đối và tính các giá trị của X, Y tương ứng, trong đó N là ký hiệu mặt ngửa xuất hiện còn S là mặt sấp.

$\mathscr{A}$	$\mathscr{B}$	$\mathscr{C}$	X	Y
N	N	N	2	3
N	N	S	2	2
N	S	N	1	2
N	S	S	1	1
S	N	N	1	2
S	N	S	1	1
S	S	N	0	1
S	S	S	0	0

Sử dụng công thức tính xác suất cổ điển (1.1a) ta có:

$$P\{X=2,Y=3\}=\frac{1}{8}\ ;\ P\{X=2,Y=2\}=\frac{1}{8}\ ;$$

$$P\{X=1,Y=2\}=\frac{2}{8}\ \ldots$$

Vậy bảng phân bố xác suất đồng thời của X và Y là:

Y / X	0	1	2	3	$\sum$
0	1/8	1/8	0	0	2/8
1	0	2/8	2/8	0	4/8
2	0	0	1/8	1/8	2/8
$\sum$	1/8	3/8	3/8	1/8	1

Phân bố xác suất của hai biến ngẫu nhiên thành phần:

Cộng các xác suất theo hàng ta được:

X	0	1	2
P	2/8	4/8	2/8

Cộng các xác suất theo cột ta được:

Y	0	1	2	3
P	1/8	3/8	3/8	1/8

Ví dụ 3.5: Có hai hộp, mỗi hộp đựng 6 bi.

Hộp I có 1 bi mang số 1, 2 bi mang số 2, 3 bi mang số 3.

Hộp II có 2 bi mang số 1, 3 bi mang số 2, 1 bi mang số 3.

Rút ngẫu nhiên từ mỗi hộp 1 bi. Gọi X,Y lần lượt là số ghi trên bi rút được từ hộp I và hộp II. Hãy lập bảng phân bố xác suất đồng thời của X,Y.

Giải: Mỗi hộp có 6 bi cho nên số các trường hợp có thể có của phép thử là $6\times 6=36$, trong đó có 2 trường hợp $(1,1)$, 3 trường hợp $(1,2)$, 4 trường hợp $(2,1)$,…

Vậy bảng phân bố xác suất đồng thời của X,Y như sau:

X \ Y	1	2	3	$\sum$
1	2/36	3/36	1/36	1/6
2	4/36	6/36	2/36	2/6
3	6/36	9/36	3/36	3/6
$\sum$	2/6	3/6	1/6	1

Ví dụ 3.6: (*Phân bố đa thức, multinomial*) Véc tơ ngẫu nhiên k chiều $X=(X_1,X_2,...,X_k)$ được gọi là có phân bố đa thức với các tham số $(n;p_1,...,p_k)$ ký hiệu $X\sim \text{MUT}(n;p_1,...,p_k)$ nếu:

$$P\{X_1=m_1,...,X_k=m_k\}=\frac{n!}{m_1!m_2!...m_k!}p_1^{m_1}p_2^{m_2}\cdots p_k^{m_k} \quad (3.13)$$

trong đó:

$$0\le m_i\le n,\ i=1,...,n;\ m_1+...+m_k=n;\ p_i>0,\ i=1,...,n;\quad p_1+\cdots+p_k=1.$$

Trường hợp $n=2$ ta được phân bố nhị thức.

Xét n phép thử độc lập, thuần nhất mỗi phép thử có k kết quả; giả sử xác suất xuất hiện kết quả thứ i là p_i thì $p_1 + ... + p_k = 1$.

Gọi X_i là số thành công của kết quả thứ i trong n phép thử thì $X = (X_1, X_2, ..., X_k)$ có phân bố đa thức $X \sim \text{MUT}(n; p_1, ..., p_k)$.

Gieo một con xúc sắc cân đối 10 lần. Tính các xác suất:

1) Có đúng 3 lần xuất hiện mặt 5 chấm (biến cố A).

2) Có 2 lần xuất hiện mặt 1 chấm, 4 lần mặt 3 chấm, 1 lần mặt 4 chấm và 3 lần mặt 6 chấm (biến cố B).

Giải: 1) Xét phép thử Bernoullivới thành công của mỗi lần thử là xuất hiện mặt có 5 chấm, vậy xác suất thành công mỗi lần thử là 1/6. Gọi X là số lần xuất hiện mặt 5 trong 10 lần thử thì X có phân bố nhị thức tham số $(10; 1/6)$, do đó:

$$P(A) = P\{X = 3\} = C_{10}^3 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^7 = 0{,}155$$

2) Gọi X_i là số lần xuất hiện mặt i chấm trong 10 phép thử thì $(X_1, X_2, X_3, X_4, X_5, X_6)$ có phân bố đa thức MUT (10; 1/6, 1/6, 1/6, 1/6, 1/6, 1/6).

$$P(B) = P\{X_1 = 2, X_2 = 0, X_3 = 4, X_4 = 1, X_5 = 0, X_6 = 3\}$$

$$= \frac{10!}{2!0!4!1!0!3!}\left(\frac{1}{6}\right)^{10} \approx 0{,}0002$$

3.3 HÀM MẬT ĐỘ XÁC SUẤT CỦA VÉC TƠ NGẪU NHIÊN LIÊN TỤC

3.3.1 Hàm mật độ xác suất đồng thời

Định nghĩa 3.3: *Hàm mật độ xác suất của véc tơ ngẫu nhiên liên tục* $(X_1, X_2, ..., X_n)$ *là hàm* n *biến* $f(x_1, x_2, ..., x_n) \geq 0$ *thoả mãn:*

$$F_{X_1X_2...X_n}(x_1,x_2,...,x_n) = \int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n} f_{X_1X_2...X_n}(t_1,t_2,...,t_n)\,dt_1dt_2\,...\,dt_n. \quad (3.14)$$

$f_{X_1...X_n}(x_1,...,x_n)$ *còn được gọi là hàm mật độ xác suất đồng thời của* các biến ngẫu nhiên $X_1,X_2,...,X_n$.

Tính chất 3.1: Để đơn giản cho cách biểu diễn ta xét trường hợp véc tơ ngẫu nhiên hai chiều (X,Y) có hàm mật độ xác suất $f_{XY}(x,y)$.

1) $f_{XY}(x,y)\geq 0$ với mọi (x,y) và $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = 1$. (3.15)

2) $P\{(X,Y)\in A\} = \iint\limits_{(x,y)\in A\cap R_{XY}} f_{XY}(x,y)dxdy$ với mọi $A\subset \mathbb{R}^2$ (3.16)

R_{XY} là miền giá trị của (X,Y).

3) $$f_{XY}(x,y) = \begin{cases} \dfrac{\partial^2}{\partial x\partial y}F_{XY}(x,y) & \text{nếu tồn tại đạo hàm tại } (x,y) \\ 0 & \text{nếu ngược lại} \end{cases} \quad (3.17)$$

3.3.2 Hàm mật độ xác suất biên

Từ bảng phân bố xác suất đồng thời của véc tơ ngẫu nhiên rời rạc (X,Y), nếu ta cộng giá trị của hàm khối lượng xác suất theo cột tức là cộng tất cả các giá trị của hàm khối lượng xác suất theo mọi giá trị của X thì ta được giá trị của hàm khối lượng xác suất với giá trị tương ứng của Y, tương tự nếu ta cộng các giá trị của hàm khối lượng xác suất theo hàng ta được giá trị của hàm khối lượng xác suất với giá trị tương ứng của X.

Trường hợp véc tơ ngẫu nhiên (X,Y) liên tục thay vì cộng ta thực hiện lấy tích phân của hàm mật độ đồng thời $f_{XY}(x,y)$ theo biến y ta được hàm mật độ của biến ngẫu nhiên thành phần X, cụ thể:

$\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = f_X(x)$ hàm mật độ xác suất của biến ngẫu nhiên X. (3.18)

$\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = f_Y(y)$ hàm mật độ xác suất của biến ngẫu nhiên Y. (3.19)

Ví dụ 3.7: Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{XY}(x,y) = \begin{cases} k & \text{nếu } 0 < y \le x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

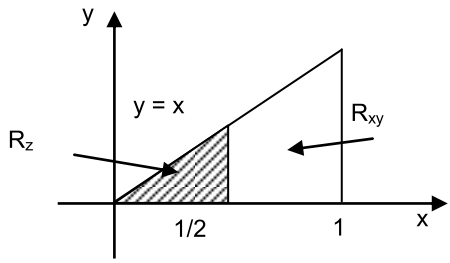
a) Tìm k.

b) Tìm các hàm mật độ xác suất của các biến ngẫu nhiên X, Y.

c) Tìm xác suất $P\{0 < X \le 1/2; 0 < Y \le 1/2\}$.

Giải: a) Miền giá trị của X, Y là tam giác R_{XY} (hình 3.1). Do đó:

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy = k\,\mathrm{dt}(R_{XY}) = k\frac{1}{2} \Rightarrow k = 2$$



Hình 3.1

b) $$f_X(x) = \begin{cases} \int_0^x 2dy = 2x & \text{nếu } 0 < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

$$f_Y(y)=\begin{cases}\int\limits_y^1 2dx=2(1-y) & \text{nếu } 0<y<1\\ 0 & \text{nếu ngược lại}\end{cases}$$

c) $P\{0<X\le 1/2;0<Y\le 1/2\}=P\{0<X\le 1/2;0<Y<X\}$

$$=\iint\limits_{R_z} f_{XY}(x,y)dxdy=2\left(\frac{1}{8}\right)=\frac{1}{4}$$

Ví dụ 3.8: Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{XY}(x,y)=\begin{cases}k & \text{nếu } |x|+|y|\le 1\\ 0 & \text{nếu ngược lại}\end{cases}.$$

a) Tìm k.

b) Tìm các hàm mật độ xác suất của các biến ngẫu nhiên X, Y.

Giải: a) Miền D: $|x|+|y|\le 1$ đối xứng qua hai trục toạ độ Ox, Oy.

Phần của D nằm trong góc phần tư thứ nhất là tam giác vuông cân $0\le x, 0\le y;\ x+y\le 1$.

Vậy D là hình vuông có độ dài cạnh bằng $\sqrt{2}$ (hình 3.2), do đó:

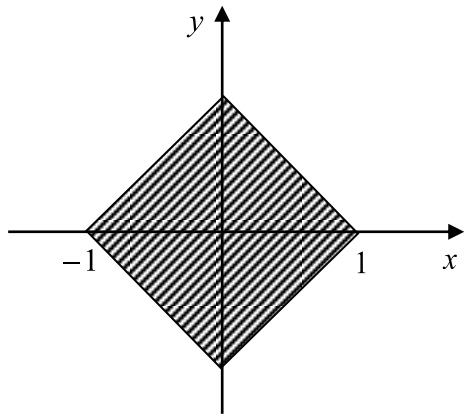
$$1=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f_{XY}(x,y)dxdy=k\text{dt D}=2k \ \Rightarrow\ k=\frac{1}{2}.$$

b) $|x|+|y|\le 1 \Leftrightarrow |x|-1\le y\le |x|+1$

$$\Leftrightarrow \begin{cases} x-1\le y\le 1-x & \text{nếu } 0\le x\le 1\\ -x-1\le y\le x+1 & \text{nếu } -1\le x\le 0\end{cases}$$

$$f_X(x)=\begin{cases}\dfrac{1}{2}\displaystyle\int_{-x-1}^{1+x} dy=1+x & \text{nếu } -1\le x\le 0\\ 0 & \text{nếu ngược lại}\\ \dfrac{1}{2}\displaystyle\int_{x-1}^{1-x} dy=1-x & \text{nếu } 0\le x\le 1\end{cases}$$

$$= \begin{cases} 1-|x| & \text{nếu } |x| \le 1 \\ 0 & \text{nếu } |x| > 1 \end{cases}$$



Hình 3.2

Do tính chất đối xứng của X và Y nên ta cũng có:

$$f_Y(y) = \begin{cases} 1-|y| & \text{nếu } |y| \le 1 \\ 0 & \text{nếu } |y| > 1 \end{cases}.$$

Ví dụ 3.9: Chọn ngẫu nhiên một điểm từ hình tròn tâm O bán kính R. Đặt X, Y là hoành độ và tung độ của điểm được chọn thì (X,Y) là véc tơ ngẫu nhiên liên tục có hàm mật độ

$$f_{XY}(x,y) = \begin{cases} k & x^2+y^2 \le R^2 \\ 0 & x^2+y^2 > R^2 \end{cases}.$$

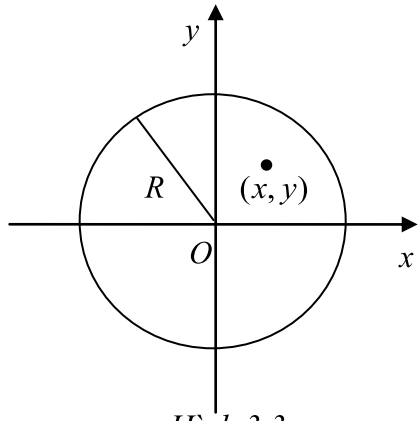
a) Xác định k.

b) Tìm các hàm mật độ thành phần $f_X(x)$, $f_Y(y)$.

c) Tìm xác suất để khoảng cách từ gốc O đến điểm được chọn không lớn hơn *a* với $a \le R$ (hình 3.3).

Giải:

a) $1 = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} f_{XY}(x,y)dxdy = k \iint\limits_{x^2+y^2\le R^2} dxdy = k\pi R^2 \quad \Rightarrow k = \dfrac{1}{\pi R^2}$



Hình 3.3

b) Khi $x^2 \le R^2$: $f_X(x) = \dfrac{1}{\pi R^2} \displaystyle\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy = \dfrac{2}{\pi R^2}\sqrt{R^2-x^2}$.

Vậy
$$f_X(x) = \begin{cases} \dfrac{2}{\pi R^2}\sqrt{R^2-x^2} & |x| \le R \\ 0 & |x| > R \end{cases}$$

Tương tự
$$f_Y(y) = \begin{cases} \dfrac{2}{\pi R^2}\sqrt{R^2-y^2} & |y| \le R \\ 0 & |y| > R \end{cases}.$$

c) Với $0 \le a \le R$:

$$P\left\{X^2+Y^2 \le a^2\right\} = \iint\limits_{x^2+y^2 \le a^2} f_{XY}(x,y)dxdy$$

$$= \frac{1}{\pi R^2} \iint\limits_{x^2+y^2 \le a^2} dxdy = \frac{\pi a^2}{\pi R^2} = \frac{a^2}{R^2}.$$

3.4 TÍNH ĐỘC LẬP CỦA CÁC BIẾN NGẪU NHIÊN

Từ định nghĩa 2.2 và các tính chất của hàm phân bố xác suất đồng thời, hàm khối lượng xác suất đồng thời, hàm mật độ xác suất đồng

thời ta có các dấu hiệu để nhận biết tính độc lập của hai biến ngẫu nhiên như sau.

Định lý 3.1: Giả sử $F_{XY}(x,y)$ là hàm phân bố xác suất của véc tơ ngẫu nhiên (X,Y). Khi đó X,Y độc lập khi và chỉ khi:

$$F_{XY}(x,y) = F_X(x)F_Y(y) \tag{3.20}$$

trong đó $F_X(x)$, $F_Y(y)$ lần lượt là hàm phân bố xác suất của X và Y.

Ví dụ 3.10: Hàm phân bố xác suất đồng thời của (X,Y) trong ví dụ 3.3 thỏa mãn (3.20) do đó độc lập.

Trường hợp véc tơ ngẫu nhiên (X,Y) rời rạc thì điều kiện (3.20) của hàm phân bố được thay bởi hàm khối lượng xác suất.

Định lý 3.2: Véc tơ ngẫu nhiên hai chiều rời rạc (X,Y), X nhận các giá trị $x_1,...,x_n$, Y nhận các giá trị $y_1,...,y_m$ với hàm khối lượng xác suất đồng thời (3.8) là độc lập khi và chỉ khi:

$$p_{XY}(x_i,y_j) = p_X(x_i)p_Y(y_j) \quad \forall i=\overline{1,n},\ j=\overline{1,m} \tag{3.21}$$

Một dấu hiệu để nhận biết hai biến ngẫu nhiên rời rạc độc lập là bảng phân bố xác suất đồng thời có tính chất:

- Hai hàng bất kỳ tỉ lệ với nhau.

- Hai cột bất kỳ tỉ lệ với nhau. (3.22)

Ví dụ 3.11: Hai biến ngẫu nhiên X và Y của ví dụ 3.4 không độc lập vì $P\{X=2\}=\frac{2}{8}$, $P\{Y=1\}=\frac{3}{8}$ nhưng $P\{X=2,Y=1\}=0$, ta cũng thấy rằng bảng phân bố xác suất đồng thời không tỉ lệ nhau. Hơn nữa ta có: $X \le Y \le X+1$

Hai biến ngẫu nhiên X và Y của ví dụ 3.5 độc lập vì thỏa mãn điều kiện (3.21), (3.22), các hàng của bảng phân bố xác suất tỉ lệ nhau theo tỉ lệ 1:2:3.

Trường hợp véc tơ ngẫu nhiên (X,Y) liên tục, bằng cách lấy đạo hàm hai vế của công thức (3.20) (xem công thức (3.17)) ta được:

Định lý 3.3: Giả sử véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất $f_{XY}(x,y)$. Khi đó X,Y độc lập khi và chỉ khi:

$$f_{XY}(x,y) = f_X(x) f_Y(y) \tag{3.23}$$

trong đó $f_X(x)$, $f_Y(y)$ lần lượt là hàm mật độ xác suất của X và Y.

Ví dụ 3.12: Véc tơ ngẫu nhiên (X,Y) của ví dụ 3.7, ví dụ 3.8, ví dụ 3.9 không độc lập vì hàm mật độ xác suất $f_{XY}(x,y) \neq f_X(x) f_Y(y)$.

Ví dụ 3.13: Véc tơ ngẫu nhiên (X,Y) có hàm mật độ:

$$f_{XY}(x,y) = \begin{cases} 4xy & \text{nếu } 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Có hai hàm mật độ thành phần:

$$f_X(x) = \begin{cases} 2x & \text{nếu } 0 < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

$$f_Y(y) = \begin{cases} 2y & \text{nếu } 0 < y < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Ta có $f_{XY}(x,y) = f_X(x) f_Y(y)$. Vậy X và Y độc lập.

Miền giá trị của véc tơ ngẫu nhiên (X,Y) có hai thành phần X, Y độc lập khi miền giá trị là hình chữ nhật có các cặp song song với trục toạ độ.

3.5 CÁC THAM SỐ ĐẶC TRƯNG CỦA VÉC TƠ NGẪU NHIÊN

3.5.1 Kỳ vọng và phương sai của các biến ngẫu nhiên thành phần

Từ bảng phân bố xác suất thành phần (3.11), (3.12) và hàm mật độ xác suất thành phần (3.18), (3.19) ta có công thức tính kỳ vọng và phương sai của biến ngẫu nhiên thành phần X, Y của véc tơ ngẫu nhiên (X,Y):

a. Trường hợp X, Y rời rạc

$$\mathrm{E}X = \sum_{i=1}^{n} x_i p_X(x_i) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i p_{XY}(x_i, y_j) \tag{3.24}$$

$$\mathrm{E}Y = \sum_{j=1}^{m} y_j p_Y(y_j) = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j p_{XY}(x_i, y_j) \tag{3.25}$$

$$\mathrm{D}X = \mathrm{E}X^2 - \big(\mathrm{E}X\big)^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i^2 p_{XY}(x_i, y_j) - \big(\mathrm{E}X\big)^2 \tag{3.26}$$

$$\mathrm{D}Y = \mathrm{E}Y^2 - \big(\mathrm{E}Y\big)^2 = \sum_{j=1}^{m}\sum_{i=1}^{n} y_j^2 p_{XY}(x_i, y_j) - \big(\mathrm{E}Y\big)^2 \tag{3.27}$$

b. Trường hợp X, Y liên tục có hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$

$$\mathrm{E}X = \int_{-\infty}^{\infty} x f_X(x)dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy \tag{3.28}$$

$$\mathrm{E}Y = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)dxdy \tag{3.29}$$

$$\mathrm{D}X = \mathrm{E}X^2 - \big(\mathrm{E}X\big)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2 f_{X,Y}(x,y)dxdy - \big(\mathrm{E}X\big)^2 \tag{3.30}$$

$$\mathrm{D}Y = \mathrm{E}Y^2 - \big(\mathrm{E}Y\big)^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2 f_{X,Y}(x,y)dxdy - \big(\mathrm{E}Y\big)^2 \tag{3.31}$$

3.5.2 Hiệp phương sai

Định nghĩa 3.4: *Hiệp phương sai (hay còn gọi là Covariance) của hai biến ngẫu nhiên X,Y ký hiệu $\mathrm{cov}(X,Y)$, là kỳ vọng toán của tích các sai lệch của hai biến ngẫu nhiên đó với kỳ vọng toán của chúng:*

$$\mathrm{cov}(X,Y) = \mathrm{E}\Big[\big(X - \mathrm{E}X\big)\big(Y - \mathrm{E}Y\big)\Big] \tag{3.32a}$$

Khai triển vế phải và áp dụng tính chất của kỳ vọng ta được

$$\mathrm{cov}(X,Y) = \mathrm{E}XY - (\mathrm{E}X)(\mathrm{E}Y) \tag{3.32b}$$

Nếu X,Y rời rạc có hàm khối lượng xác suất đồng thời theo công thức (3.8) thì:

$$\mathrm{E}XY = \sum_{j=1}^{m}\sum_{i=1}^{n} x_i y_j p_{XY}(x_i, y_j) \tag{3.33}$$

Nếu X,Y liên tục có hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$ thì:

$$\mathrm{E}\,XY = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{X,Y}(x,y)dxdy \tag{3.34}$$

Tính chất 3.2:

1) $\mathrm{cov}(X,Y) = \mathrm{cov}(Y,X)$ (3.35)

2) $\mathrm{cov}(X,X) = \mathrm{D}X$ (3.36)

3) $\mathrm{cov}(aX+c,bY+d) = ab\,\mathrm{cov}(Y,X)$ với mọi hằng số a,b,c,d. (3.37)

4) Nếu X,Y độc lập thì $\mathrm{cov}(X,Y)=0$ nhưng ngược lại chưa chắc đúng.

3.5.3 Ma trận hiệp phương sai

Định nghĩa 3.5: *Cho véc tơ ngẫu nhiên* $X=(X_1,X_2,...,X_n)$.

Ma trận $M=\left[C_{ij}\right]_{n\times n}$ *với* $C_{ij}=\mathrm{cov}(X_i,X_j)$ (3.38)

được gọi là ma trận hiệp phương sai (ma trận covariance) của véc tơ ngẫu nhiên X.

Tính chất 3.3:

1) Ma trận hiệp phương sai là ma trận đối xứng.

2) Với mọi $t_1,t_2,...,t_n \in \mathbb{R}^n$ luôn có $\sum_j\sum_i C_{ij}t_it_j \geq 0$.

3) Các định thức con chính của M không âm.

Nói cách khác ma trận hiệp phương sai M là ma trận của dạng toàn phương không âm.

3.5.4 Hệ số tương quan

Định nghĩa 3.6: *Hệ số tương quan của hai biến ngẫu nhiên* X,Y *với phương sai khác 0 được ký hiệu và định nghĩa bởi công thức:*

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{D}X}\sqrt{\mathrm{D}Y}} \tag{3.39}$$

Tính chất 3.4:

1) $-1 \le \rho_{X,Y} \le 1$ với mọi X, Y. (3.40)

2) Nếu X, Y độc lập thì $\rho_{X,Y} = 0$, nhưng điều ngược lại chưa chắc đúng.

3) Với mọi hằng số a, b, c, d

$$\rho_{aX+c,bY+d} = \begin{cases} \rho_{X,Y} & \text{nếu } ab > 0 \\ -\rho_{X,Y} & \text{nếu } ab < 0 \end{cases} \quad (3.41)$$

4) $Y = aX + b$, $a \neq 0$ khi và chỉ khi:

$$\rho_{X,Y} = \begin{cases} 1 & \text{nếu } a > 0 \\ -1 & \text{nếu } a < 0 \end{cases} \quad (3.42)$$

Ý nghĩa của hệ số tương quan

Hệ số tương quan đo mức độ phụ thuộc tuyến tính giữa X và Y. Khi $|\rho_{X,Y}|$ càng gần 1 thì tính chất quan hệ tuyến tính càng chặt, khi $|\rho_{X,Y}|$ càng gần 0 thì sự phụ thuộc tuyến tính càng ít, càng lỏng lẻo.

Khi $\rho_{X,Y} = 0$ ta nói X và Y không tương quan.

Ví dụ 3.14: Xét véc tơ ngẫu nhiên (X, Y) ở ví dụ 3.4

Y / X	0	1	2	3
0	1/8	1/8	0	0
1	0	2/8	2/8	0
2	0	0	1/8	1/8

Có bảng phân bố xác suất biên

X	0	1	2
P	2/8	4/8	2/8

Y	0	1	2	3
P	1/8	3/8	3/8	1/8

$$\mathrm{E}X = 0.\frac{2}{8} + 1.\frac{4}{8} + 2.\frac{2}{8} = 1;$$

$$\mathrm{E}X^2 = 0^2.\frac{2}{8} + 1^2.\frac{4}{8} + 2^2.\frac{2}{8} = \frac{3}{2} \Rightarrow \mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \frac{1}{2}$$

$$\mathrm{E}Y = 0.\frac{1}{8} + 1.\frac{3}{8} + 2.\frac{3}{8} + 3.\frac{1}{8} = \frac{3}{2};\ \mathrm{E}Y^2 = 3 \Rightarrow \mathrm{D}Y = \frac{3}{4}.$$

$$\mathrm{E}XY = 0.0.\frac{1}{8} + 0.1.\frac{1}{8} + 0.2.0 + 0.3.0 + 1.0.0 + 1.1.\frac{2}{8} + 1.2.\frac{2}{8} +$$

$$+ 1.3.0 + 2.0.0 + 2.1.0 + 2.2.\frac{1}{8} + 2.3.\frac{1}{8} = 2$$

Hiệp phương sai

$$\mathrm{cov}(X,Y) = \mathrm{E}XY - \mathrm{E}X\,\mathrm{E}Y = 2 - 1.\frac{3}{2} = \frac{1}{2}.$$

Hệ số tương quan

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{D}X}\sqrt{\mathrm{D}Y}} = \frac{1/2}{\sqrt{(1/2)(3/4)}} = \sqrt{\frac{2}{3}}.$$

Ma trận hiệp phương sai: $M = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 3/4 \end{bmatrix}$.

Ví dụ 3.15: Xét kênh viễn thông nhị phân được cho trong ví dụ 1.38 chương 1. Ký hiệu X là đầu vào và Y đầu ra của kênh. (X,Y) là véc tơ ngẫu nhiên thỏa mãn:

$P\{X=0\} = 0,5$; $P\{Y=1|X=0\} = 0,1$ và $P\{Y=0|X=1\} = 0,2$.

a) Tìm bảng phân bố xác suất đồng thời của (X,Y).

b) Tìm bảng phân bố xác suất thành phần X và Y. X và Y có độc lập không?

c) Tính $\mathrm{E}X$, $\mathrm{E}Y$; $\mathrm{D}X$, $\mathrm{D}Y$

d) Tìm hiệp phương sai, hệ số tương quan và ma trận hiệp phương sai.

Giải: a)

Y / X	0	1	Σ
0	0,45	0,05	0,5
1	0,10	0,40	0,5
Σ	0,55	0,45	1

b) Bảng phân bố xác suất thành phần X và Y

X	0	1
P	0,5	0,5

Y	0	1
P	0,55	0,45

Hai hàng của bảng phân bố xác suất đồng thời của (X,Y) không tỉ lệ, do đó X, Y không độc lập.

Ta cũng thấy rằng:

$$P\{X=0;Y=0\}=0,45\neq P\{X=0\}P\{Y=0\}$$
$$=0,55\times 0,5=0,275.$$

c) Áp dụng các công thức (3.24)-(3.27) ta có:

$$\mathrm{E}X=0,5,\ \mathrm{E}X^2=0,5\ \Rightarrow \mathrm{D}X=0,25$$

$$\mathrm{E}Y=0,45,\ \mathrm{E}Y^2=0,45\ \Rightarrow \mathrm{D}Y=0,2475.$$

d) Áp dụng công thức (3.32), (3.33), (3.38), (3.39) ta được:

Hiệp phương sai:

$$\mathrm{E}XY=0,4\Rightarrow \mathrm{cov}(X,Y)=0,4-0,5\times 0,45=0,175.$$

Hệ số tương quan:

$$\rho_{X,Y}=\frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{D}X}\sqrt{\mathrm{D}Y}}=\frac{0,175}{\sqrt{0,25\times 0,2475}}=0,704.$$

Ma trận hiệp phương sai: $M = \begin{bmatrix} 0,25 & 0,175 \\ 0,175 & 0,2475 \end{bmatrix}$.

Ta thấy giá trị $\rho_{X,Y} = 0,704$ khá xa 1, do đó Y không phụ thuộc tuyến tính đối với X.

3.6 HÀM CỦA CÁC BIẾN NGẪU NHIÊN

3.6.1 Hàm của một biến ngẫu nhiên

Giả sử X là một biến ngẫu nhiên và $g(x)$ là hàm liên tục một biến số. Khi đó:

$$Y = g(X) \tag{3.43}$$

cũng là một biến ngẫu nhiên và gọi là hàm của biến ngẫu nhiên X.

Điều kiện liên tục của hàm $g(x)$ có thể thay bởi điều kiện đo được. Tuy nhiên vấn đề này vượt khỏi nội dung của giáo trình.

3.6.1.1 Trường hợp rời rạc

Nếu X là biến ngẫu nhiên rời rạc có bảng phân bố xác suất:

X	x_1	x_2	...
P	$P_X(x_1)$	$P_X(x_2)$	...

thì cũng có bảng xác định giá trị và xác suất tương ứng của Y:

Y	$g(x_1)$	$g(x_2)$	...
P	$p_X(x_1)$	$p_X(x_2)$	...

Chú ý rằng biến ngẫu nhiên X nhận các giá trị x_i khác nhau nhưng các giá trị tương ứng $g(x_i)$ của biến ngẫu nhiên Y có thể trùng nhau, vì vậy bảng ghi giá trị của Y theo các giá trị $g(x_i)$ và xác suất tương ứng ở trên có thể chưa phải là bảng phân bố xác suất của Y. Tuy nhiên từ bảng trên ta có thể suy ra bảng phân bố xác suất của Y.

3.6.1.2 Trường hợp liên tục

Cho X liên tục có hàm phân bố xác suất $F_X(x)$ và hàm mật độ xác suất $f_X(x)$, bằng phương pháp giải tích ta có thể tìm được hàm phân bố xác suất và hàm mật độ xác suất của biến ngẫu nhiên $Y = g(X)$.

- Trường hợp $g(x)$ là một song ánh. Khi đó $g(x)$ hoặc đơn điệu tăng hoặc đơn điệu giảm:

+ $g(x)$ đơn điệu tăng

$$F_Y(y) = P\{g(X) \le y\} = P\left\{X \le g^{-1}(y)\right\}$$

$$= F_X\left(g^{-1}(y)\right) \Rightarrow F_X(x) = F_Y\left(g(x)\right)$$

$$f_X(x) = f_Y\left(g(x)\right)g'(x)$$

hay $$f_Y(y) = \frac{f_X\left(g^{-1}(y)\right)}{g'\left(g^{-1}(y)\right)} = \frac{f_X\left(g^{-1}(y)\right)}{\left|g'\left(g^{-1}(y)\right)\right|} = f_X(x)\left|\frac{dx}{dy}\right|. \qquad (3.44a)$$

+ $g(x)$ đơn điệu giảm

$$F_Y(y) = P\{g(X) \le y\} = P\left\{X > g^{-1}(y)\right\}$$

$$= 1 - F_X\left(g^{-1}(y)\right) \Rightarrow F_X(x) = 1 - F_Y\left(g(x)\right)$$

$$f_X(x) = -f_Y\left(g(x)\right)g'(x)$$

hay $$f_Y(y) = -\frac{f_X\left(g^{-1}(y)\right)}{g'\left(g^{-1}(y)\right)} = \frac{f_X\left(g^{-1}(y)\right)}{\left|g'\left(g^{-1}(y)\right)\right|} = f_X(x)\left|\frac{dx}{dy}\right|. \qquad (3.44b)$$

- Trường hợp $g(x)$ là một hàm liên tục bất kỳ (không đơn điệu) ta chia miền xác định của $g(x)$ thành các miền mà $g(x)$ đơn điệu và áp dụng công thức (3.44a) hoặc (3.44b) trong các miền này. Ngoài ra cũng có thể tính trực tiếp xác suất $P\{Y = g(X) \le y\}$, từ đó suy ra hàm phân bố xác suất và hàm mật độ xác suất của Y.

Ví dụ 3.16: Cho $g_1(x)=x^3$, $g_2(x)=x^2$, $g_3(x)=e^{-x}$. Giả sử X là một biến ngẫu nhiên liên tục có hàm phân bố xác suất $F_X(x)$ và hàm mật độ xác suất $f_X(x)$.

a) Tìm hàm mật độ xác suất của $Y=g_1(X)$.

b) Tìm hàm mật độ xác suất của $Z=g_2(X)$.

c) Tìm hàm mật độ xác suất của $T=g_3(X)$.

d) Tìm hàm mật độ xác suất $f_Y(y)$ và hàm phân bố xác suất $F_Y(y)$. Tính xác suất biến cố $P\{2,5<Y<5\}$ nếu hàm mật độ xác suất $f_X(x)$ của X xác định bởi:

$$f_X(x)=\begin{cases}\dfrac{2}{x^2} & \text{nếu } 1\le x\le 2\\ 0 & \text{nếu ngược lại}\end{cases}$$

Giải: a) $y=g_1(x)=x^3$ đơn điệu tăng; $x=\sqrt[3]{y}$, $\dfrac{dx}{dy}=\dfrac{1}{3\sqrt[3]{y^2}}$. Áp dụng công thức (3.44a) ta có:

$$f_Y(y)=f_X(x)\frac{dx}{dy}=\frac{1}{3}\frac{f_X(\sqrt[3]{y})}{\sqrt[3]{y^2}}.$$

b) $g_2(x)=x^2$ không đơn điệu. Ta tính trực tiếp $P\{Z\le z\}$ như sau:

- Với $z<0$ thì $P\{Z\le z\}=0$ vì $Z=X^2$ không nhận giá trị âm.

- Với $z\ge 0$ ta có:

$$F_Z(z)=P\{Z\le z\}=P\{X^2\le z\}=P\{-\sqrt{z}\le X\le\sqrt{z}\}$$
$$=F_X(\sqrt{z})-F_X(-\sqrt{z}).$$

Vậy hàm mật độ xác suất của Z:

$$f_Z(z)=\begin{cases}\dfrac{1}{2\sqrt{z}}\left(f_X(\sqrt{z})+f_X(-\sqrt{z})\right) & \text{nếu } z>0\\ 0 & \text{nếu } z\le 0\end{cases}$$

c) $g_3(x) = e^{-x}$ đơn điệu giảm; $t = e^{-x} \Leftrightarrow x = -\ln t$; $\dfrac{dx}{dt} = -\dfrac{1}{t}$.

Áp dụng công thức (3.44b) ta có: $f_T(t) = f_X(x)\left|\dfrac{dx}{dt}\right| = f_X(-\ln t)\dfrac{1}{t}$.

Ta cũng có thể tìm $f_T(t)$ trực tiếp bằng cách tính $P\{T < t\}$

Với $t > 0$ ta có:

$$F_T(t) = P\{T \le t\} = P\left\{e^{-X} \le t\right\} = P\{X > -\ln t\} = 1 - F_X(-\ln t)$$

Khi $t \le 0$ biến cố $\{T \le t\}$ là biến cố không thể vì $T = e^{-X}$ chỉ nhận giá trị dương, do đó $P\{T \le t\} = 0$. Vậy hàm mật độ xác suất của T xác định như sau:

$$f_T(t) = \begin{cases} \dfrac{f_X(-\ln t)}{t} & \text{nếu} \quad t > 0 \\ 0 & \text{nếu} \quad t \le 0 \end{cases}$$

d) Thay giả thiết vào kết quả ở mục a) ta được:

$$f_Y(y) = \frac{1}{3}\frac{f_X(\sqrt[3]{y})}{\sqrt[3]{y^2}} = \begin{cases} \dfrac{2}{3\sqrt[3]{y^4}} & \text{nếu} \quad 1 \le y \le 8 \\ 0 & \text{nếu ngược lại} \end{cases}$$

$$F_Y(y) = \begin{cases} 0 & \text{nếu} \quad y \le 1 \\ 2\left(1 - \dfrac{1}{\sqrt[3]{y}}\right) & \text{nếu} \quad 1 < y \le 8 \\ 1 & \text{nếu} \quad y > 8 \end{cases}$$

$$P\{2,5 < Y < 5\} = P\left\{2,5 < X^3 < 5\right\} = P\left\{\sqrt[3]{2,5} < X < \sqrt[3]{5}\right\}$$

$$= \int_{\sqrt[3]{2,5}}^{\sqrt[3]{5}} \frac{2}{x^2}\,dx = -\frac{2}{x}\Big|_{\sqrt[3]{2,5}}^{\sqrt[3]{5}} = 2\left(\frac{1}{\sqrt[3]{2,5}} - \frac{1}{\sqrt[3]{5}}\right)$$

hoặc $$P\{2,5 < Y < 5\} = F_Y(5) - F_Y(2,5) = 2\left(\frac{1}{\sqrt[3]{2,5}} - \frac{1}{\sqrt[3]{5}}\right).$$

Ví dụ 3.17: Cho X là biến ngẫu nhiên với hàm phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$. Tìm hàm mật độ xác suất của $Y = aX + b$; a, b là hai hằng số thực, $a \neq 0$.

Tìm a, b sao cho $Y = aX + b$ có phân bố chuẩn tắc (xem công thức (2.33) chương 2).

Giải: Xét hàm số $y = g(x) = ax + b$, áp dụng công thức (3.44a), (3.44b) ta được:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

$$= \frac{1}{\sqrt{2\pi}|a|\sigma} \exp\left\{-\frac{1}{2\sigma^2}\left[\left(\frac{y-b}{a}\right)-\mu\right]^2\right\}$$

$$= \frac{1}{\sqrt{2\pi}|a|\sigma} \exp\left\{-\frac{1}{2a^2\sigma^2}\left[y-(a\mu+b)\right]^2\right\}$$

Vậy Y có phân bố chuẩn $\mathbf{N}(a\mu+b;a^2\sigma^2)$.

Chọn $a = 1/\sigma$, $b = -\mu/\sigma$ thì Y có phân bố chuẩn tắc. Vậy:

Nếu $X \sim \mathbf{N}(\mu;\sigma^2)$ thì $Y = \dfrac{X-\mu}{\sigma} \sim \mathbf{N}(0;1)$.

3.6.2 Hàm của hai biến ngẫu nhiên

Giả sử X, Y là hai biến ngẫu nhiên và $g(x,y)$ là một hàm hai biến liên tục cho trước, khi đó $Z = g(X,Y)$ là một biến ngẫu nhiên được gọi là hàm của hai biến ngẫu nhiên X và Y.

Ta có thể xác định hàm khối lượng xác suất hoặc hàm mật độ xác suất của Z qua hàm khối lượng xác suất đồng thời hoặc hàm mật độ xác suất đồng thời của X, Y.

3.6.2.1 Trường hợp rời rạc

Nếu X, Y có tập các giá trị $\{x_1, x_2, ..., x_m\}$ và $\{y_1, y_2, ..., y_n\}$ thì $Z = g(X,Y)$ có tập các giá trị là $\left\{g(x_i, y_j)\middle| i = 1,...m; j = 1,...,n\right\}$.

Định lý 3.4: Giả sử $Z = g(X,Y)$. Khi đó:

$$P\{Z = z\} = \sum_{\{(i,j)|g(x_i,y_j)=z\}} P\{X = x_i; Y = y_j\} \quad (3.45)$$

Ví dụ 3.18: Xét $Z = X + Y$, với véc tơ ngẫu nhiên (X,Y) cho trong hai trường hợp sau:

a) Ví dụ 3.2

X	0	0	0	0	1	1	1	1	2	2	2	2
Y	0	1	2	3	0	1	2	3	0	1	2	3
X+Y	0	1	2	3	1	2	3	4	2	3	4	5
P	1/8	1/8	0	0	0	2/8	2/8	0	0	0	1/8	1/8

Vậy bảng phân bố xác suất của $Z = X + Y$ xác định như sau:

Z = X+Y	0	1	2	3	4	5
P	1/8	1/8	2/8	2/8	1/8	1/8

b) Ví dụ 3.3

X	1	1	1	2	2	2	3	3	3
Y	1	2	3	1	2	3	1	2	3
X+Y	2	3	4	3	4	5	4	5	6
P	2/36	3/36	1/36	4/36	6/36	2/36	6/36	9/36	3/36

Vậy bảng phân bố xác suất của $Z = X + Y$ xác định như sau:

Z = X+Y	2	3	4	5	6
P	2/36	7/36	13/36	11/36	3/36

3.6.2.2 Trường hợp liên tục

Giả sử véc tơ ngẫu nhiên (X,Y) liên tục có hàm mật độ xác suất đồng thời $f_{XY}(x,y)$. $Z = g(X,Y)$ sẽ là biến ngẫu nhiên liên tục.

Với mỗi $z \in 3$, ký hiệu $D_z = \{(x,y) \in R_{XY} \,|\, g(x,y) \le z\}$, (3.46)

R_{XY} là tập giá trị của (X,Y).

Định lý 3.5: Hàm phân bố xác suất của biến ngẫu nhiên $Z = g(X,Y)$ được xác định như sau:

$$F_Z(z) = \iint_{D_z} f_{XY}(x,y)dxdy \tag{3.47}$$

3.6.3 Hàm phân bố của tổng hai biến ngẫu nhiên

Định lý 3.6: Giả sử $Z = X + Y$ là tổng của hai biến ngẫu nhiên X, Y.

(i) Nếu X, Y rời rạc thì hàm khối lượng xác suất của Z xác định như sau:

$$\begin{aligned} P\{Z = z\} &= \sum_{x \in R_X} P\{X = x; Y = z - x\} \\ &= \sum_{y \in R_Y} P\{Y = y; X = z - y\} \end{aligned} \tag{3.48}$$

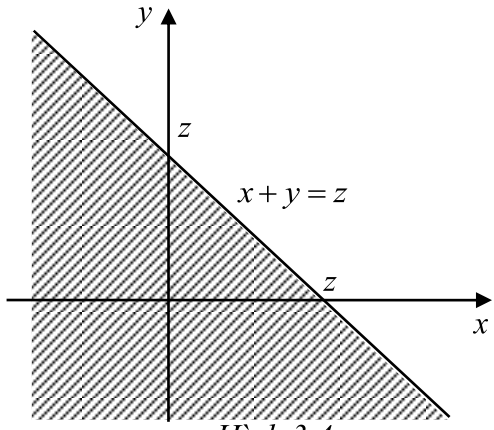
(ii) Nếu X, Y liên tục có hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$ thì hàm mật độ xác suất của Z:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx\,. \tag{3.49}$$

Chứng minh: *(i)* Áp dụng công thức (3.45) ta có :

$$\begin{aligned} P\{Z = z\} &= \sum_{\{(i,j): x_i + y_j = z\}} P\{X = x_i; Y = y_j\} = \sum_{x_i} P\{X = x_i; Y = z - x_i\} \\ &= \sum_{y_j \in R_Y} P\{X = z - y_j; Y = y_j\} \end{aligned}$$

(ii) Miền D_z trong công thức (3.46) là tập các điểm $(x,y) \in R_{XY}$ nằm dưới đường thẳng $x + y = z$.



Hình 3.4

Do đó: $F_Z(z) = P\{X+Y<z\} = \int\limits_{-\infty}^{\infty}\left[\int\limits_{-\infty}^{z-x} f_{XY}(x,y)dy\right]dx$

Vậy:

$$f_Z(z) = \frac{d}{dz}F_Z(z) = \int\limits_{-\infty}^{\infty}\left[\frac{\partial}{\partial z}\int\limits_{-\infty}^{z-x} f_{XY}(x,y)dy\right]dx = \int\limits_{-\infty}^{\infty} f_{XY}(x,z-x)dx$$

Hệ quả 3.1: Giả sử $Z = X+Y$ là tổng của hai biến ngẫu nhiên độc lập X, Y.

(i) Nếu X, Y rời rạc thì hàm khối lượng xác suất của Z có dạng

$$\begin{aligned} P\{Z=z\} &= \sum_{x\in R_X} P\{X=x\}P\{Y=z-x\} \\ &= \sum_{y\in R_Y} P\{Y=y\}P\{X=z-y\} \end{aligned} \tag{3.50}$$

Tổng $\sum\limits_{x\in R_X}$ lấy theo x chạy trong miền giá trị của X và đồng thời $z-x$ thuộc miền giá trị của Y. Tương tự tổng $\sum\limits_{y\in R_Y}$ lấy theo y chạy trong miền giá trị của Y và đồng thời $z-y$ thuộc miền giá trị của X.

(ii) Nếu X, Y liên tục có hàm mật độ xác suất đồng thời $f_X(x)f_Y(y)$ thì hàm mật độ xác suất của Z có dạng:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx = f_X(z) * f_Y(z) \qquad (3.51a)$$

Như vậy biến ngẫu nhiên là tổng hai biến ngẫu nhiên độc lập liên tục X, Y có hàm mật độ xác suất là tích chập của hai hàm mật độ $f_X(z) * f_Y(z)$.

Vì tích chập có tính giao hoán nên ta cũng nhận được công thức:

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y)dy = f_Y(z) * f_X(z) \qquad (3.51b)$$

Ví dụ 3.19: Giả sử X, Y là hai biến ngẫu nhiên độc lập có phân bố Poisson với các tham số λ_1, λ_2. Xét biến ngẫu nhiên $Z = X + Y$. Áp dụng công thức (3.50) ta có:

$$P\{Z=n\} = \sum_{k=0}^{n} P\{X=k\}P\{Y=n-k\} = \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}$$

$$= e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n$$

Vậy $Z = X + Y$ cũng có phân bố Poisson với tham số $\lambda_1 + \lambda_2$.

Ví dụ 3.20: Giả sử X, Y là hai biến ngẫu nhiên độc lập có phân bố đều trong khoảng $(0;1)$. Xét biến ngẫu nhiên $Z = X + Y$.

Hàm mật độ xác suất của X và Y lần lượt là:

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}; \quad f_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Miền giá trị của X và Y là khoảng $(0;1)$, do đó miền giá trị của $Z = X + Y$ là khoảng $(0;2)$ và $R_{XY} = (0;1)\times(0;1)$.

Áp dụng công thức (4.51a) ta có:

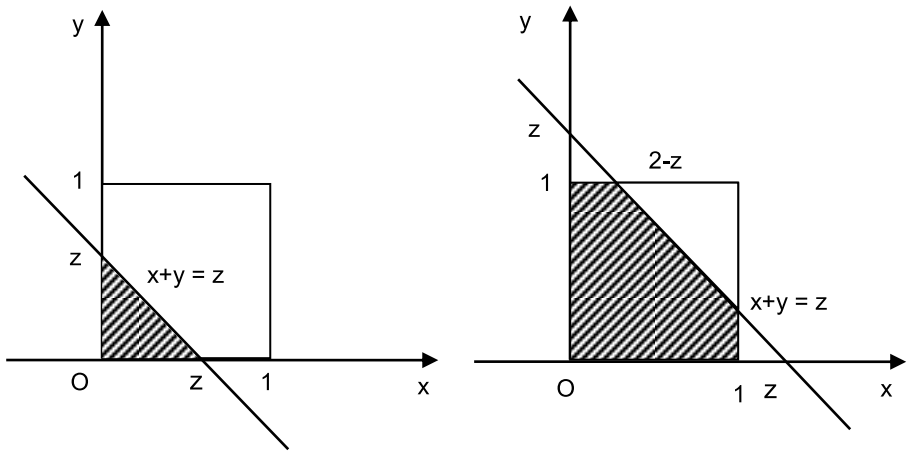
$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx$$

$$= \int_0^z f_X(x) f_Y(z-x)dx = \begin{cases} \int_0^z dx & 0 < z < 1 \\ \int_{z-1}^1 dx & 1 < z < 2 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Vậy: $$f_Z(z) = \begin{cases} z & 0 < z < 1 \\ 2-z & 1 < z < 2 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Ta cũng có thể tính trực tiếp bằng cách xét $F_Z(z) = P\{X + Y \le z\}$

Hàm mật độ xác suất đồng thời của X và Y (xem hình 3.5):



Hình 3.5

$$f_{XY}(x,y)=\begin{cases}1 & 0<x<1, 0<y<1\\ 0 & \text{nếu ngược lại}\end{cases}$$

$$F_Z(z)=P\{X+Y\le z\}=\iint\limits_{x+y\le z} f_{XY}(x,y)dxdy$$

$$=\iint\limits_{x+y\le z;(x,y)\in R_{XY}} dxdy = \text{ diện tích miền gạch chéo.}$$

- Trường hợp $0<z<1$:

$$F_Z(z)=\iint\limits_{x+y\le z;(x,y)\in R_{XY}} dxdy = \text{diện tích miền gạch chéo} = \frac{z^2}{2},$$

$$f_Z(z)=\frac{d}{dz}F_Z(z)=z.$$

- Trường hợp $1<z<2$:

$$F_Z(z)=\iint\limits_{x+y\le z;(x,y)\in R_{XY}} dxdy = \text{diện tích miền gạch chéo.}$$

$$=1-\frac{(2-z)^2}{2},\ f_Z(z)=\frac{d}{dz}F_Z(z)=2-z.$$

3.6.4 Hai hàm của hai biến ngẫu nhiên liên tục

Giả sử X, Y là hai biến ngẫu nhiên và $z=g(x,y),\ w=h(x,y)$ là hai hàm hai biến liên tục cho trước. Khi đó ta có thể thiết lập một cặp biến ngẫu nhiên mới Z, W bằng công thức:

$$\begin{cases}Z=g(X,Y)\\ W=h(X,Y)\end{cases} \tag{3.52}$$

Với mỗi $(z,w)\in 3^2$, ký hiệu:

$$D_{zw}=\{(x,y)\in R_{XY}\,|\,g(x,y)\le z; h(x,y)\le w\}$$

trong đó R_{XY} là miền giá trị của (X,Y). Ta có:

$$\{Z\le z; W\le w\}=\{g(X,Y)\le z; h(X,Y)\le w\}=\{(X,Y)\in D_{zw}\} \tag{3.53}$$

Áp dụng công thức (3.16) đối với véc tơ ngẫu nhiên liên tục (X,Y) có hàm mật độ $f_{XY}(x,y)$ ta được:

$$F_{ZW}(z,w) = \iint\limits_{D_{zw}} f_{XY}(x,y)dxdy \qquad (3.54)$$

The công thức (3.54) ta có thể xác định hàm mật độ $f_{ZW}(z,w)$ từ hàm mật độ $f_{XY}(xy,)$ như sau:

Giả sử T là ánh xạ từ $\mathbb{R}^2$ vào $\mathbb{R}^2$ xác định bởi:

$$T(x,y) = \big(g(x,y), h(x,y)\big)$$

Giả thiết T là ánh xạ 1 - 1 (đơn ánh) từ miền $R \subset \mathbb{R}^2$ lên miền $\Delta \subset \mathbb{R}^2$. Khi đó T có ánh xạ ngược T^{-1}: $\Delta \to R$ xác định như sau:

Với mỗi $(z,w) \in \Delta$ tồn tại duy nhất $(x,y) \in R$ sao cho $T(x,y) = (z,w)$, nghĩa là:

$$\begin{cases} z = g(x,y) \\ w = h(x,y) \end{cases}$$

ta ký hiệu $T^{-1}(z,w) = (x,y)$.

Định thức Jacobi của T^{-1} được xác định theo công thức sau:

$$J(z,w) = \frac{D(x,y)}{D(z,w)} = \begin{vmatrix} \dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w} \\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w} \end{vmatrix}$$

Ta có định thức Jacobi của T và công thức liên hệ với định thức Jacobi của T^{-1}:

$$J(x,y) = \frac{D(z,w)}{D(x,y)} = \begin{vmatrix} \dfrac{\partial z}{\partial x} & \dfrac{\partial z}{\partial y} \\ \dfrac{\partial w}{\partial x} & \dfrac{\partial w}{\partial y} \end{vmatrix} = \frac{1}{J(z,w)}.$$

Định lý 3.7: Giả sử X, Y là hai biến ngẫu nhiên liên tục với hàm mật độ xác suất đồng thời $f_{XY}(x,y)$. Ký hiệu R_{XY} là miền giá trị của (X,Y):

$$R_{XY} = \left\{(x,y) \in \mathbb{R}^2 \,\middle|\, f_{X,Y}(x,y) > 0\right\}$$

Giả sử $T(x,y) = \big(z(x,y), w(x,y)\big)$ là ánh xạ 1 - 1 khả vi liên tục từ miền $R_{XY} \subset \mathbb{R}^2$ lên miền $\Delta \subset \mathbb{R}^2$. Khi đó $Z = z(X,Y), W = w(X,Y)$ là hai biến ngẫu nhiên có hàm mật độ xác suất đồng thời

$$f_{ZW}(z,w) = \begin{cases} f_{XY}\big(x(z,w), y(z,w)\big)\left|\dfrac{D(x,y)}{D(z,w)}\right| & \text{nếu} \quad (z,w) \in \Delta \\ 0 & \text{nếu ngược lại} \end{cases} \qquad (3.55)$$

Ví dụ 3.21: Xét phép biến đổi T từ $\mathbb{R}^2$ vào $\mathbb{R}^2$ xác định bởi: $T(x,y) = \big(z(x,y), w(x,y)\big)$ trong đó:

$$\begin{cases} z = x + y \\ w = x \end{cases} \Leftrightarrow \begin{cases} x = w \\ y = z - w \end{cases}$$

$$\frac{D(x,y)}{D(z,w)} = \begin{vmatrix} \dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w} \\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1 .$$

Áp dụng công thức (3.55) ta nhận được:

$$f_{ZW}(z,w) = f_{XY}(w, z-w) .$$

Từ công thức (3.18) ta nhận được công thức (3.49) của biến ngẫu nhiên $Z = X + Y$:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{ZW}(z,w)dw = \int_{-\infty}^{\infty} f_{XY}(w, z-w)dw$$

$$= \int_{-\infty}^{\infty} f_{XY}(x, z-x)dx.$$

Ví dụ 3.22: Xét phép biến đổi T từ $\mathbb{R}^2$ vào $\mathbb{R}^2$ xác định bởi:

$$T(x,y)=\big(z(x,y),w(x,y)\big) \text{ trong đó: } \begin{cases} z=xy \\ w=x \end{cases} \Leftrightarrow \begin{cases} x=w \\ y=\dfrac{z}{w} \end{cases}$$

$$\frac{D(x,y)}{D(z,w)}=\begin{vmatrix} \dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w} \\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w} \end{vmatrix}=\begin{vmatrix} 0 & 1 \\ \dfrac{1}{w} & -\dfrac{z}{w^2} \end{vmatrix}=-\frac{1}{w}.$$

Áp dụng công thức (3.55) ta nhận được:

$$f_{ZW}(z,w)=\left|\frac{1}{w}\right| f_{XY}\left(w,\frac{z}{w}\right).$$

Từ công thức (3.18) ta nhận được công thức hàm mật độ của biến ngẫu nhiên $Z=XY$:

$$f_Z(z)=\int_{-\infty}^{\infty} f_{ZW}(z,w)dw=\int_{-\infty}^{\infty}\left|\frac{1}{w}\right| f_{XY}\left(w,z/w\right)dw$$

$$f_Z(z)=\int_{-\infty}^{\infty} f_{XY}\left(x,z/x\right)\frac{dx}{|x|} \tag{3.56}$$

Ví dụ 3.23: Giả sử X, Y là hai biến ngẫu nhiên độc lập có phân bố đều trong khoảng $(0;1)$. Hàm mật độ xác suất đồng thời của X và Y:

$$f_{XY}(x,y)=\begin{cases} 1 & 0<x<1, 0<y<1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Miền giá trị của $Z=XY$ cũng là khoảng $(0;1)$, do đó:

$$f_{XY}\left(w,\frac{z}{w}\right)=\begin{cases} 1 & 0<w<1, 0<z/w<1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

$$f_{XY}\left(w,\frac{z}{w}\right)=\begin{cases} 1 & 0<z<w<1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

$$f_{ZW}(z,w)=\left|\frac{1}{w}\right| f_{XY}\left(w,\frac{z}{w}\right)=\begin{cases}\dfrac{1}{w} & 0<z<w<1\\ 0 & \text{nếu ngược lại}\end{cases}$$

$$f_Z(z)=\int_{-\infty}^{\infty} f_{ZW}(z,w)dw=\int_{z}^{1}\frac{1}{w}dw=-\ln z\,,\quad 0<z<1$$

Vậy: $f_Z(z)=\begin{cases}-\ln z & 0<z<1\\ 0 & \text{nếu ngược lại}\end{cases}$

Ví dụ 3.24: Xét phép biến đổi T xác định bởi:

$$T(x,y)=\big(z(x,y),w(x,y)\big),\ \text{trong đó}\ \begin{cases}z=x/y\\ w=y\end{cases}$$

$$T^{-1}(z,w)=(x,y),\ \begin{cases}x=zw\\ y=w\end{cases}$$

$$\frac{D(x,y))}{D(z,w}=\begin{vmatrix}\dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w}\\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w}\end{vmatrix}=\begin{vmatrix}w & z\\ 0 & 1\end{vmatrix}=w\,.$$

Áp dụng công thức (3.55) ta nhận được:

$$f_{ZW}(z,w)=|w|\,f_{XY}(zw,w)\,.$$

Từ công thức (3.18) ta nhận được công thức hàm mật độ của biến ngẫu nhiên $Z=\dfrac{X}{Y}$:

$$f_Z(z)=\int_{-\infty}^{\infty} f_{ZW}(z,w)dw=\int_{-\infty}^{\infty}|w|\,f_{XY}(zw,w)dw \qquad (3.57)$$

Ví dụ 3.25: Giả sử X, Y là hai biến ngẫu nhiên độc lập có phân bố chuẩn tắc $\mathbf{N}(0;1)$.

Xét $Z=\dfrac{X}{Y}$. Áp dụng công thức (3.57) và từ tính chất độc lập của X, Y ta được.

$$f_Z(z) = \int_{-\infty}^{\infty} |w| f_{XY}(zw, w)\, dw$$

$$= \int_{-\infty}^{\infty} |w| f_X(zw) f_Y(w)\, dw = \int_{-\infty}^{\infty} |w| \frac{1}{2\pi} e^{-w^2(1+z^2)/2} dw$$

$$= \frac{1}{\pi} \int_0^{\infty} w e^{-w^2(1+z^2)/2} dw = \frac{1}{\pi(1+z^2)}, \quad -\infty < z < \infty.$$

Vậy $Z = \dfrac{X}{Y}$ là biến ngẫu nhiên có phân bố Cauchy tham số 1.

Ví dụ 3.26: Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ đồng thời $f_{XY}(x, y)$.

Xét phép biến đổi $\begin{cases} x = r\cos\theta \\ y = r\sin\theta; \ r \geq 0, 0 \leq \theta < 2\pi \end{cases}$

Định thức Jacobi: $\dfrac{D(x,y)}{D(r,\theta)} = \begin{vmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r.$

Đặt:

$$\begin{cases} X = R\cos\Theta \\ Y = R\sin\Theta \end{cases} \tag{3.58}$$

Áp dụng công thức (3.55) ta được:

$$f_{R\Theta}(r, \theta) = r f_{XY}(r\cos\theta, r\sin\theta) \tag{3.59}$$

3.6.5 Kỳ vọng của hàm các biến ngẫu nhiên

3.6.5.1 Kỳ vọng của hàm một biến ngẫu nhiên

Định lý 3.8: Kỳ vọng của $Y = g(X)$ được tính theo công thức sau (xem công thức 2.49):

$$\mathrm{E}Y = \begin{cases} \sum\limits_i g(x_i)p_X(x_i) & \text{nếu } X \text{ rời rạc có } p_X(x_i) = P\left\{X = x_i\right\} \\ \int\limits_{-\infty}^{\infty} g(x)f_X(x)dx & \text{nếu } X \text{ liên tục có hàm mật độ } f_X(x) \end{cases} \quad (3.60)$$

Chứng minh: Nếu X rời rạc có miền giá trị $\left\{x_i \middle| i \in I\right\}$ thì Y nhận các giá trị có dạng $g(x_i)$.

Ký hiệu $p_X(x_i) = P\left\{X = x_i\right\}$ là hàm khối lượng xác suất của X.

Đặt $I = \bigcup\limits_k I_k$, trong đó các tập I_k thỏa mãn:

- $i, j \in I_k$ thì $g(x_i) = g(x_j)$,

- $i \in I_k$, $j \in I_{k'}$, $k \neq k'$ thì $g(x_i) \neq g(x_j)$.

Nói cách khác các tập I_k tạo thành một phân hoạch của I theo các giá trị bằng nhau của miền giá trị của $Y = g(X)$

Vậy: $\mathrm{E}Y = \sum\limits_k g(x_k) \sum\limits_{i \in I_k} p_X(x_i) = \sum\limits_{i \in I} g(x_i)p_X(x_i)$.

Trường hợp X liên tục, $g(x)$ đơn điệu, áp dụng công thức (3.43), (3.44a), (3.44b) ta có

$$\mathrm{E}Y = \int\limits_{-\infty}^{\infty} yf_Y(y)dy = \int\limits_{-\infty}^{\infty} g(x)f_Y(g(x))\left|g'(x)\right|dx = \int\limits_{-\infty}^{\infty} g(x)f_X(x)dx.$$

Nếu $g(x)$ không đơn điệu, ta phân hoạch miền lấy tích phân thành các miền đơn điệu của $g(x)$ và nhận được công thức trên.

Định lý 3.8 cũng còn đúng đối với trường hợp hàm của n biến được phát biểu như sau.

3.6.5.2 Kỳ vọng của hàm n biến ngẫu nhiên

Định lý 3.9: Giả sử $X_1, ..., X_n$ là n biến ngẫu nhiên. Xét $Y = g(X_1, ..., X_n)$, kỳ vọng của Y được tính theo công thức sau (xem công thức 2.49):

- Trường hợp $X_1,...,X_n$ rời rạc với hàm khối lượng xác suất $p_{X_1...X_n}(x_{i_1},....,x_{i_n})$ thì:

$$\mathrm{E}\,Y=\sum_{i_1}...\sum_{i_n}g(x_{i_1},...,x_{i_n})\,p_{X_1...X_n}(x_{i_1},....,x_{i_n}) \tag{3.61a}$$

- Trường hợp $X_1,..., X_n$ liên tục với hàm mật độ $f_{X_1...X_n}(x_1,...,x_n)$ thì:

$$\mathrm{E}\,Y=\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}g(x_1,...,x_n)\,f_{X_1...X_n}(x_1,...,x_n)dx_1...dx_n \tag{3.61b}$$

3.6.5.3 Tính chất tuyến tính của kỳ vọng, kỳ vọng của các biến ngẫu nhiên độc lập

a) Với mọi biến ngẫu nhiên $X_1,...,X_n$, với mọi hằng số $a_1,...,a_n$:

$$\mathrm{E}\left(\sum_{k=1}^{n}a_kX_k\right)=\sum_{k=1}^{n}a_k\,\mathrm{E}\left(X_k\right) \tag{3.62}$$

b) Nếu X,Y là hai biến ngẫu nhiên độc lập; g, h là hai hàm liên tục tuỳ ý thì:

$$\mathrm{E}\left[g(X)h(Y)\right]=\mathrm{E}\left[g(X)\right]\mathrm{E}\left[h(Y)\right] \tag{3.63}$$

Tính chất này có thể mở rộng cho n biến ngẫu nhiên độc lập $X_1,...,X_n$.

$$\mathrm{E}\left[\prod_{k=1}^{n}g_k(X_k)\right]=\prod_{k=1}^{n}\mathrm{E}\left[g_k(X_k)\right] \tag{3.64}$$

Chứng minh: a) Ta chứng minh công thức đối với trường hợp hai biến ngẫu nhiên X,Y liên tục. Trường hợp khác được chứng minh tương tự.

Áp dụng công thức (3.61b) ta có:

$$\begin{aligned}\mathrm{E}\left(aX+bY\right)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)\,f_{XY}(x,y)dxdy\\&=a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{XY}(x,y)dxdy+b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf_{XY}(x,y)dxdy\end{aligned}$$

$$= a\int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f_{XY}(x,y)dy\right]dx + b\int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty} f_{XY}(x,y)dx\right]dy$$

$$= a\int_{-\infty}^{\infty} xf_X(x)dx + b\int_{-\infty}^{\infty} yf_Y(y)dy = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y).$$

b) X, Y độc lập do đó $f_{XY}(x,y) = f_X(x)f_Y(y)$.

$$\mathrm{E}[g(X)h(Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y)dxdy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy$$

$$= \int_{-\infty}^{\infty} g(x)f_X(x)dx\int_{-\infty}^{\infty} h(y)f_Y(y)dy = \mathrm{E}[g(X)]\mathrm{E}[h(Y)].$$

3.7 PHÂN BỐ CÓ ĐIỀU KIỆN VÀ KỲ VỌNG CÓ ĐIỀU KIỆN

3.7.1 Phân bố có điều kiện và kỳ vọng có điều kiện của biến ngẫu nhiên rời rạc

Định nghĩa 3.7: *Giả sử X biến ngẫu nhiên rời rạc nhận các giá trị $\{x_1, x_2, ..., x_n\}$ và B là một biến cố trong cùng phép thử với X có xác suất $P(B) > 0$. Khi đó bảng phân bố xác suất của X với điều kiện B được xác định như sau:*

X/B	x_1	x_2	...	x_I	...	x_N
P	$p_{X\|B}(x_1\|B)$	$p_{X\|B}(x_2\|B)$	...	$p_{X\|B}(x_i\|B)$	...	$p_{X\|B}(x_n\|B)$

trong đó:

$$p_{X|B}(x_i|B) = \frac{P(\{X = x_i\}B)}{P(B)};\ i = \overline{1,n}. \qquad (3.65)$$

Hàm $p_{X|B}(x_i|B)$ xác định bởi công thức (3.65) được gọi là *hàm khối lượng xác suất với điều kiện B*.

Định nghĩa 3.8: *Kỳ vọng của X với điều kiện B được định nghĩa và ký hiệu:*

$$\mathrm{E}\left[X\middle|B\right]=\sum_{i=1}^{n}x_i p_{X|B}(x_i|B) \tag{3.66}$$

Định nghĩa 3.9: *Giả sử X, Y có tập các giá trị lần lượt là $\{x_1,x_2,...,x_n\}$ và $\{y_1,y_2,...,y_m\}$. Với mỗi y_j ta có bảng phân bố xác suất có điều kiện của X với điều kiện biến cố $\{Y=y_j\}$:*

$X\|Y=y_j$	x_1	x_2	…	x_i	…	x_n
P	$p_{X\|Y}(x_1\|y_j)$	$p_{X\|Y}(x_2\|y_j)$	…	$p_{X\|Y}(x_i\|y_j)$	…	$p_{X\|Y}(x_n\|y_j)$

trong đó:

$$p_{X|Y}(x_i\,|\,y_j)=\frac{p_{XY}(x_i,y_j)}{p_Y(y_j)}\;;\; i=\overline{1,n},\;\; j=\overline{1,m}\;. \tag{3.67}$$

Tương tự với mỗi x_i ta có bảng phân bố xác suất có điều kiện của Y với điều kiện $\{X=x_i\}$:

$Y\|X=x_i$	y_1	y_2	…	y_j	…	y_m
P	$p_{Y\|X}(y_1\|x_i)$	$p_{Y\|X}(y_2\|x_i)$	…	$p_{Y\|X}(y_j\|x_i)$	…	$p_{Y\|X}(y_m\|x_i)$

trong đó:

$$p_{Y|X}(y_j\,|\,x_i)=\frac{p_{XY}(x_i,y_j)}{p_X(x_i)}\;;\; i=\overline{1,n},\;\; j=\overline{1,m}\;. \tag{3.68}$$

Tính chất 3.5:

a) $0\le p_{X|Y}(x_i\,|\,y_j)\le 1\;;\;\forall i=\overline{1,n},\;\forall j=\overline{1,m}\;.$

b) Với mỗi j: $\sum_{i=1}^{n}p_{X|Y}(x_i\,|\,y_j)=1.$

c) Nếu X, Y độc lập thì $p_{X|Y}(x_i\,|\,y_j)=p_X(x_i)$ và:

$$p_{Y|X}(y_j \mid x_i) = p_Y(y_j).$$

Định nghĩa 3.10: *Kỳ vọng có điều kiện của* X *với điều kiện* $Y = y_j$ *và kỳ vọng có điều kiện của* Y *với điều kiện* $X = x_i$ *tương ứng được tính theo công thức sau:*

$$\begin{aligned} \mathrm{E}\left[X \middle| Y = y_j\right] &= \sum_{i=1}^{n} x_i p(x_i \middle| Y = y_j); \\ \mathrm{E}\left[Y \middle| X = x_i\right] &= \sum_{j=1}^{m} y_j p(y_j \middle| X = x_i). \end{aligned} \tag{3.69}$$

Ví dụ 3.27: Thực hiện lặp lại nhiều lần phép thử Bernoulli, xác suất xuất hiện của biến cố A trong mỗi lần thử là p, $0 < p < 1$. Gọi Y là biến ngẫu nhiên chỉ lần thử đầu tiên xuất hiện biến cố A. Gọi B là biến cố: "Trong n lần thử đầu tiên có duy nhất một lần xuất hiện biến cố A".

a) Tìm bảng phân bố xác suất của Y.

b) Tìm phân bố của Y với điều kiện B.

Giải: a) Ta có bảng phân bố xác suất của Y (xem ví dụ 2.9)

Y	1	2	…	k	…
P	p	qp	…	$q^{k-1}p$	…

với $q = 1 - p$.

b) Phân bố của Y với điều kiện B:

Khi $P(B) > 0$ thì $P\left(\{Y = k\} \middle| B\right) = \dfrac{P\left(\{Y = k\}\, B\right)}{P(B)}$.

- Rõ ràng khi $k > n$ thì biến cố $\{Y = k\}$ kéo theo trong n phép thử đầu tiên biến cố A không xuất hiện. Do đó $\{Y = k\}\, B = \varnothing$, vậy $P\left(\{Y = k\}\, B\right) = 0$.

- Khi $k \le n$, áp dụng công thức Bernoulli ta có:

$$P(B) = C_n^1 pq^{n-1} = npq^{n-1}.$$

Mặt khác:

$$P\big(\{Y=k\}\,B\big)=P\{\text{chỉ xuất hiện biến cố } A \text{ ở lần thử thứ } k\}=pq^{n-1}$$

Vậy: $$P\big(\{Y=k\}\big|B\big)=\begin{cases}\dfrac{1}{n} & \text{nếu} \quad k\le n\\ 0 & \text{nếu} \quad k>n\end{cases}.$$

Ví dụ 3.28: Thống kê dân cư của một thành phố nọ ở độ tuổi trưởng thành về thu nhập hàng tháng X và lứa tuổi Y thu được kết quả trong bảng sau:

Y / X	30	45	70
1	0,01	0,02	0,05
2	0,03	0,06	0,10
3	0,18	0,21	0,15
4	0,07	0,08	0,04

trong đó: $X=1, 2, 3, 4$ tương ứng chỉ thu nhập triệu đồng /tháng.

$Y=30, 45, 70$ chỉ độ tuổi của người dân trong khoảng: 25÷35, 35÷55, 55÷85.

Tìm thu nhập trung bình theo lứa tuổi.

Giải: Thu nhập trung bình theo lứa tuổi là kỳ vọng có điều kiện của X theo Y.

Với $Y=30$ bảng phân bố xác suất điều kiện tương ứng:

X/Y = 30	1	2	3	4
P	0,01/0,29	0,03/0,29	0,18/0,29	0,07/0,29

Từ đó: $\mathrm{E}\big[X\big|Y=30\big]=1\cdot\dfrac{1}{29}+2\cdot\dfrac{3}{29}+3\cdot\dfrac{18}{29}+4\cdot\dfrac{7}{29}=\dfrac{89}{29}=3,069$.

Tương tự:

$$\mathrm{E}\big[X\big|Y=45\big]=\frac{109}{37}=2,9459.\quad \mathrm{E}\big[X\big|Y=70\big]=\frac{86}{34}=2,5294.$$

Vậy thu nhập trung bình ở độ tuổi 30 là 3.069.000đ/tháng, độ tuổi 45 là 2.945.900đ/tháng và độ tuổi 70 là 2.529.400 đ/tháng.

3.7.2 Phân bố có điều kiện và kỳ vọng có điều kiện của biến ngẫu nhiên liên tục

Định nghĩa 3.11: *Xét biến ngẫu nhiên X và biến cố B trong cùng một phép thử. Giả sử $P(B)>0$. Hàm phân bố xác suất của X với điều kiện B được định nghĩa và ký hiệu như sau:*

$$F_{X|B}(x\,|\,B)=P\{X\le x|\,B\}=\frac{P\big(\{X\le x\}\,B\big)}{P(B)}. \tag{3.70}$$

Hàm mật độ xác suất của X với điều kiện B được định nghĩa và ký hiệu như sau:

$$f_{X|B}(x\,|\,B)=\frac{dF_{X|B}(x\,|\,B)}{dx}. \tag{3.71}$$

Định nghĩa 3.12: *Giả sử $f_{XY}(x,y)$ là hàm mật độ xác suất đồng thời của hai biến ngẫu nhiên liên tục X, Y và $f_X(x)$ là hàm mật độ của biến ngẫu nhiên thành phần X. Hàm phân bố xác suất của Y với điều kiện $\{X=x\}$ được định nghĩa và ký hiệu như sau:*

$$F_{Y|X}(y\,|\,x)=\int_{-\infty}^{y}\frac{f_{XY}(x,v)}{f_X(x)}dv\text{, nếu }\ f_X(x)>0. \tag{3.72}$$

Đạo hàm của hàm phân bố xác suất có điều kiện $\dfrac{d}{dy}F_{Y|X}(y\,|\,x)$ được gọi là hàm mật độ xác suất có điều kiện của Y với điều kiện $\{X=x\}$ và ký hiệu là $f_{Y|X}(y\,|\,x)$. (Đây là hàm của biến y còn x đóng vai trò là tham số).

$$f_{Y|X}(y\,|\,x)=\frac{f_{XY}(x,y)}{f_X(x)}$$

với mọi $y\in\mathbb{R}$ và x thoả mãn $f_X(x)>0$. (3.73)

Tương tự ta có hàm phân bố xác suất và hàm mật độ xác suất có điều kiện của X với $\{Y = y\}$

$$F_{X|Y}(x \mid y) = \int_{-\infty}^{x} \frac{f_{XY}(u,y)}{f_Y(y)} du \text{, nếu } f_Y(y) > 0. \qquad (3.74)$$

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} \text{ với mọi } x \in \mathbb{R} \text{ và } y \text{ thoả mãn } f_Y(y) > 0. \qquad (3.75)$$

Tính chất 3.6:

a) $f_{X|Y}(x \mid y) \geq 0$.

b) Với mỗi y thỏa mãn $f_Y(y) > 0$ thì có $\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)dx = 1$.

c) Nếu X, Y độc lập thì $f_{X|Y}(x \mid y) = f_X(x)$ và $f_{Y|X}(y \mid x) = f_Y(y)$.

Ví dụ 3.29: Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời đã cho trong ví dụ 3.8 thì:

$$f_{X|Y}(x \mid y) = \begin{cases} \dfrac{1}{2(1-|y|)} & \text{nếu } |y| \leq 1 \text{ và } |x|+|y| \leq 1 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

Định nghĩa 3.13: *Giả sử hai biến ngẫu nhiên X, Y có $f_{Y|X}(y \mid x)$ là hàm mật độ xác suất có điều kiện của Y với điều kiện $\{X = x\}$. Khi đó kỳ vọng của Y với điều kiện $\{X = x\}$ được ký hiệu và định nghĩa theo công thức sau:*

$$\mathrm{E}\left[Y \middle| X = x\right] = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x)dy. \qquad (3.76)$$

$\mathrm{E}\left[Y \middle| X = x\right]$ xác định ở công thức (3.69), (3.76) là một hàm của x, được gọi là hàm hồi qui của Y đối với X.

$\mathrm{E}\left[X \middle| Y = y\right]$ là một hàm của y, được gọi là hàm hồi qui của X đối với Y.

Ví dụ 3.30: Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời đã cho trong ví dụ 3.7.

$$f_{XY}(x,y)=\begin{cases}2 & \text{nếu } 0<y\le x<1\\ 0 & \text{nếu ngược lại}\end{cases}$$

Tìm hàm mật độ có điều kiện $f_{X|Y}(x\,|\,y)$, $f_{Y|X}(y\,|\,x)$ và kỳ vọng điều kiện.

Giải: Theo kết quả của ví dụ 3.7 ta có:

$$f_X(x)=2x \qquad \text{nếu } 0<x<1$$

$$f_Y(y)=2(1-y) \qquad \text{nếu } 0<y<1$$

Áp dụng công thức (3.73), (3.77) ta được:

$$f_{Y|X}(y\,|\,x)=\frac{1}{x} \qquad y\le x<1,\ 0<x<1$$

$$f_{X|Y}(x\,|\,y)=\frac{1}{1-y} \qquad y\le x<1,\ 0<x<1.$$

Kỳ vọng với điều kiện $\{X=x\}$

$$\mathrm{E}\big[Y\big|X=x\big]=\int_{-\infty}^{\infty} yf_{Y|X}(y\,|\,x)dy=\int_0^x y\left(\frac{1}{x}\right)dy=\left.\frac{y^2}{2x}\right|_0^x=\frac{x}{2}$$

nếu $0<x<1$.

Kỳ vọng với điều kiện $\{Y=y\}$

$$\mathrm{E}\big[X\big|Y=y\big]=\int_{-\infty}^{\infty} xf_{X|Y}(x\,|\,y)dx$$

$$=\int_y^1 x\left(\frac{1}{1-y}\right)dx=\left.\frac{x^2}{2(1-y)}\right|_{x=y}^{x=1}=\frac{1+y}{2} \text{ nếu } 0<y<1.$$

3.7.3 Biến ngẫu nhiên kỳ vọng có điều kiện

Hàm hồi qui của Y đối với X là một hàm phụ thuộc biến x: $H(x)=\mathrm{E}\big(Y\big|X=x\big)$.

Xét biến ngẫu nhiên của hàm $H(x)$ xác định bởi công thức (3.43):

$$\mathrm{E}\left[Y\middle|X\right]=H(X),\text{ trong đó }H(x)=\mathrm{E}\left[Y\middle|X=x\right] \qquad (3.77)$$

Tương tự nếu đặt hàm hồi qui của X đối với Y là $G(y)=\mathrm{E}\left[X\middle|Y=y\right]$, ta định nghĩa:

$$\mathrm{E}\left[X\,|\,Y\right]=G(Y),\text{ trong đó: }G(y)=\mathrm{E}\left[X\middle|Y=y\right] \qquad (3.78)$$

Biến ngẫu nhiên xác định bởi công thức (3.77), (3.78) gọi là biến ngẫu nhiên kỳ vọng có điều kiện.

Áp dụng công thức (3.60) ta được $\mathrm{E}\left[Y\right]=\mathrm{E}\left[\mathrm{E}\left(Y\,|\,X\right)\right]$, cụ thể:

- Nếu X rời rạc với hàm khối lượng xác suất $p_X(x_i)=P\{X=x_i\}$ thì:

$$\mathrm{E}\left[Y\right]=\sum_i \mathrm{E}\left[Y\middle|X=x_i\right]p_X(x_i). \qquad (3.79)$$

- Nếu X liên tục với hàm mật độ xác suất $f_X(x)$ thì:

$$\mathrm{E}\left[Y\right]=\int_{-\infty}^{\infty}\mathrm{E}\left[Y\middle|X=x\right]f_X(x)dx. \qquad (3.80)$$

3.8 PHÂN BỐ CHUẨN NHIỀU CHIỀU

3.8.1 Khái niệm phân bố chuẩn *n* chiều

Định nghĩa 3.14: *Véc tơ ngẫu nhiên* $(X_1,X_2,...,X_n)$ *được gọi là có phân bố chuẩn* n *chiều nếu có mật độ đồng thời*

$$f(x)=\frac{1}{\sqrt{(2\pi)^n\det M}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})M^{-1}(\boldsymbol{x}-\boldsymbol{a})^t\right\} \qquad (3.81)$$

trong đó $\boldsymbol{x}=(x_1,x_2,...,x_n)\in\mathbb{R}^n$; $\boldsymbol{a}=(\mu_1,\mu_2,...,\mu_n)\in\mathbb{R}^n$; $M=\left[C_{ij}\right]_{n\times n}$ là ma trận đối xứng xác định dương; ma trận cột $(\boldsymbol{x}-\boldsymbol{a})^t$ là chuyển vị của ma trận hàng $(\boldsymbol{x}-\boldsymbol{a})$.

Ký hiệu $(X_1, X_2, ..., X_n) \sim \mathbf{N}(\mathbf{a}, M)$

Định nghĩa 3.15: *Hàm đặc trưng của véc tơ ngẫu nhiên* $(X_1, X_2, ..., X_n)$ *là hàm của* n *biến thực* $\boldsymbol{t} = (t_1, t_2, ..., t_n)$ *xác định như sau:*

$$\varphi_{X_1 ... X_n}(t_1, t_2, ..., t_n) = \mathrm{E}\left[e^{i(t_1 X_1 + \cdots + t_n X_n)} \right] \quad (3.82)$$

Định lý 3.10: Nếu véc tơ ngẫu nhiên $X = (X_1, X_2, ..., X_n)$ có phân bố chuẩn n chiều với mật độ đồng thời xác định như trong định nghĩa trên thì:

a) Hàm đặc trưng:

$$\varphi_X(\boldsymbol{t}) = \exp\left(i\langle \boldsymbol{t}, \boldsymbol{a} \rangle - \frac{1}{2} \langle \boldsymbol{t}, M^{-1} \boldsymbol{t}^t \rangle \right); \ \boldsymbol{t} = (t_1, t_2, ..., t_n) \quad (3.83)$$

trong đó $\langle \boldsymbol{t}, \boldsymbol{a} \rangle = t_1 \mu_1 + t_2 \mu_2 + \cdots + t_n \mu_n$ là tích vô hướng của hai véc tơ $\boldsymbol{t}$ và $\boldsymbol{a}$.

b) Với mọi $i = 1, ..., n$: $\mu_i = \mathrm{E}X_i$.

c) M là ma trận hiệp phương sai của véc tơ ngẫu nhiên $X = (X_1, X_2, ..., X_n)$:

$$C_{ij} = \mathrm{cov}(X_i, X_j). \quad (3.84)$$

d) Các biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập khi và chỉ khi chúng không tương quan, nghĩa là M là ma trận chéo.

e) Biến ngẫu nhiên thành phần X_i có phân bố chuẩn $\mathbf{N}(\mu_i; C_{ii})$.

f) $(Y_1, Y_2, ..., Y_n) = H(X_1, X_2, ..., X_n)^t$ cũng có phân bố chuẩn, trong đó H là ma trận vuông cấp n bất kỳ có $\det H \neq 0$.

Ví dụ 3.31: Tìm hàm phân bố của biến ngẫu nhiên $T = 2X - 3Y + 7Z$ biết rằng:

$$(X, Y, Z) \sim \mathbf{N}(\mathbf{a}, M), \ \mathbf{a} = (8, 3, -2)$$

$$M = \begin{bmatrix} 3 & -0,5 & 0,4 \\ -0,5 & 2 & 0,7 \\ 0,4 & 0,7 & 5 \end{bmatrix}$$

Giải: Theo định lý trên T có phân bố chuẩn, ta chỉ cần tìm kỳ vọng và phương sai.

$$\mathrm{E}T = 2\,\mathrm{E}X - 3\,\mathrm{E}Y + 7\,\mathrm{E}Z = (2)(8) - (3)(3) + (7)(-2) = -7$$

$$\mathrm{D}T = \mathrm{E}\big[T - \mathrm{E}T\big]^2 = \mathrm{E}\big[2(X-8) - 3(Y-3) + 7(Z+2)\big]^2$$

$$= 4\,\mathrm{E}\big[(X-8)\big]^2 + 9\,\mathrm{E}\big[(Y-3)\big]^2 + 49\,\mathrm{E}\big[(Z+2)\big]^2$$
$$-12\,\mathrm{E}\big[(X-8)(Y-3)\big] + 28\,\mathrm{E}\big[(X-8)(Z+2)\big] - 42\,\mathrm{E}\big[(Y-3)(Z+2)\big]$$

$$= (4)(3) + (9)(2) + (49)(5) - (12)(-0,5)$$
$$+ (28)(0,4) - (42)(0,7) = 262,8$$

Vậy $T \sim \mathbf{N}(-7; 262,8)$.

3.8.2 Phân bố chuẩn hai chiều

Định nghĩa 3.16: *Véc tơ ngẫu nhiên (X,Y) được gọi là có phân bố chuẩn hai chiều nếu hàm mật độ xác suất đồng thời có dạng:*

* Nếu $\rho^2 \neq 1$:

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}Q(x,y)\right). \qquad (3.85)$$

trong đó:

$$Q(x,y) = \frac{1}{(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}.$$

* Nếu $\rho^2 = 1$:

$$f_X(x,y) = \frac{1}{2\pi\delta_1\delta_2} \exp\left\{-\frac{1}{2}Q(x,y)\right\}$$

$$Q(x,y)=\left(\frac{x-\mu_1}{\delta_1}\right)^2+\left(\frac{y-\mu_2}{\delta_2}\right)^2.$$

Định lý 3.11: Giả sử véc tơ ngẫu nhiên (X,Y) có phân bố chuẩn hai chiều với hàm mật độ xác suất đồng thời xác định như trong định nghĩa 3.13. Khi đó:

a) X có phân bố chuẩn $\mathbf{N}(\mu_1;\sigma_1^2)$, Y có phân bố chuẩn $\mathbf{N}(\mu_2;\sigma_2^2)$.

b) Hệ số tương quan $\rho(X,Y)=\rho$.

c) X, Y không tương quan khi và chỉ khi X, Y độc lập.

d) Hàm mật độ xác suất của X với điều kiện $Y=y$:

$$f_{X|Y}(x\,|\,y)=\frac{f_{XY}(x,y)}{f_Y(y)}=\frac{1}{\sigma_1\sqrt{1-\rho^2}}\varphi\left(\frac{x-\mu_1-(\sigma_1/\sigma_2)\rho(y-\mu_2)}{\sigma_1\sqrt{1-\rho^2}}\right) \quad (3.86)$$

trong đó $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$ là hàm mật độ xác suất của biến ngẫu nhiên chuẩn tắc $\mathbf{N}(0;1)$.

Nhận xét 3.1:

a) Ma trận tương quan của (X,Y) là :

* Nếu $\rho^2\neq 1$:

$$M=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\Rightarrow M^{-1}=\frac{1}{1-\rho^2}\begin{bmatrix}1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2)\\ -\rho/(\sigma_1\sigma_2) & 1/\sigma_2^2\end{bmatrix} \quad (3.87)$$

* Nếu $\rho^2=1$:

$$\mu=\begin{bmatrix}\delta_1^{-2} & 0\\ 0 & \delta_2^2\end{bmatrix}\Rightarrow\mu^{-1}=\begin{bmatrix}\frac{1}{\delta_1^{-2}} & 0\\ 0 & \frac{1}{\delta_2^2}\end{bmatrix}$$

b) Từ định lý 3.11 suy ra rằng phân bố của X với điều kiện $Y = y$ có phân bố chuẩn với kỳ vọng $\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)$ và phương sai $\sigma_1^2(1-\rho^2)$ (xem công thức 3.86), do đó kỳ vọng có điều kiện:

$$\mathrm{E}\left[X \middle| Y = y\right] = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)\,. \tag{3.88}$$

là một hàm tuyến tính theo y. Đây là một tính chất rất quan trọng của phân bố chuẩn 2 chiều.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

3.1 Bảng phân bố xác suất của X và Y cho phép xác định bảng phân bố xác suất đồng thời của (X,Y).

Đúng ☐ Sai ☐.

3.2 Bảng phân bố xác suất đồng thời của (X,Y) xác định bảng phân bố xác suất của hai biến ngẫu nhiên thành phần X và Y.

Đúng ☐ Sai ☐.

3.3 Nếu hai biến ngẫu nhiên X, Y độc lập thì bảng phân bố xác suất của X và Y cho phép xác định bảng phân bố xác suất đồng thời của (X,Y).

Đúng ☐ Sai ☐.

3.4 Hai biến ngẫu nhiên độc lập có hiệp phương sai bằng 0.

Đúng ☐ Sai ☐.

3.5 Hai biến ngẫu nhiên có hiệp phương sai bằng 0 thì độc lập.

Đúng ☐ Sai ☐.

3.6 Hiệp phương sai luôn nhận giá trị dương.

Đúng ☐ Sai ☐.

3.7 Nếu $\mathrm{Y} = \mathrm{aX} + \mathrm{b},\ \mathrm{a} \neq 0$ thì hệ số tương quan $\rho_{X,Y}$ luôn luôn bằng 1.

Đúng ☐ Sai ☐.

3.8 Nếu $\{x_1, x_2, ..., x_n\}$ là tập giá trị của X thì $\{f(x_1), f(x_2), ..., f(x_n)\}$ là tập giá trị của hàm hồi quy $f(x) = E\left[Y \middle| X = x\right]$ của Y đối với X.

Đúng ☐ Sai ☐.

3.9 Nếu hai biến ngẫu nhiên X, Y độc lập thì hàm hồi quy $\mathrm{f(x)} = \mathrm{E}\left[\mathrm{Y} \middle| \mathrm{X} = \mathrm{x}\right]$ của Y đối với X và hàm hồi quy $\mathrm{g(y)} = \mathrm{E}\left[\mathrm{X} \middle| \mathrm{Y} = \mathrm{y}\right]$ của X đối với Y là hai hàm hằng.

Đúng ☐ Sai ☐.

3.10 Nếu hiệp phương sai của hai biến ngẫu nhiên bằng 0 thì hai kỳ vọng của chúng bằng nhau ($\text{cov}(X,Y)=0$ thì $EX=EY$).

Đúng ☐ Sai ☐ .

3.11 Hàm mật độ xác suất đồng thời của véc tơ ngẫu nhiên liên tục *(X,Y)* bằng tích của hai hàm mật độ xác suất thành phần X và Y.

Đúng ☐ Sai ☐ .

3.12 Giả sử (X,Y) là véc tơ ngẫu nhiên phân bố chuẩn, khi đó X, Y độc lập khi và chỉ khi X, Y không tương quan.

Đúng ☐ Sai ☐ .

3.13 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời như sau:

Y / X	y_1	y_2	y_3
x_1	0,18	0,22	0,16
x_2	0,08	0,16	0,20

Tìm bảng phân bố xác suất của hai biến ngẫu nhiên thành phần *X, Y.*

3.14 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời như sau:

Y / X	-1	1
-1	1/6	1/4
0	1/6	1/8
1	1/6	1/8

Hãy xác định EX, EY, $\text{cov}(X,Y)$ và $\rho_{X,Y}$.

3.15 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời như sau:

Y / X	-1	0	1
-1	4/15	1/15	4/15
0	1/15	2/15	1/15
1	0	2/15	0

Hãy xác định *EX, EY*, $\text{cov}(X,Y)$ và $\rho_{X,Y}$. X, Y có độc lập không?

3.16 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời như sau:

Y / X	1	2	3
1	0,12	0,15	0,03
2	0,28	0,35	0,07

a) Chứng minh rằng X, Y có độc lập.

b) Tìm quy luật phân bố của biến ngẫu nhiên $Z = XY$.

c) Tính các kỳ vọng $\mathrm{E}X, \mathrm{E}Y, \mathrm{E}Z$.

3.17 Cho X, Y là hai biến ngẫu nhiên độc lập có bảng phân bố xác suất như sau:

X	0	1	2	3
P	0,4	0,3	0,2	0,1

Y	0	1	2	3	4
P	0,1	0,3	0,4	0,15	0,05

Tìm bảng phân bố xác suất đồng thời của X, Y. Tính xác suất $P\{X > Y\}$.

3.18 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời như sau:

Y / X	1	2	3
1	0,15	0,20	0,10
2	0,35	0,05	0,15

Hai biến ngẫu nhiên X, Y có độc lập không. Tính xác suất $P\{X = 1 | Y = 2\}$.

3.19 Gieo đồng thời một con xúc sắc và một đồng tiền. Gọi X là biến ngẫu nhiên chỉ số chấm của con xúc sắc và Y là biến ngẫu nhiên chỉ mặt sấp (1) hay mặt ngửa (0) của đồng tiền. Lập bảng phân bố xác suất đồng thời của X và Y.

3.20 Cho bảng phân bố xác suất đồng thời của X, Y

X \ Y	26	30	41	50
23	0,05	0,08	0,12	0,04
27	0,09	0,30	0,11	0,21

Tìm bảng phân bố xác suất điều kiện của Y khi $X = 26$ và của X khi $Y = 27$.

3.21 Cho bảng phân bố xác suất đồng thời của X, Y

X \ Y	1	3	4	8
3	0,15	0,06	0,25	0,04
6	0,30	0,10	0,03	0,07

a) Tìm kỳ vọng có điều kiện của Y khi $X = 1$.

b) Tìm các kỳ vọng EX, EY và phương sai DX, DY.

3.22 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời:

X \ Y	-1	0	1
-1	4α	α	4α
0	α	2α	α
1	0	2α	0

a) Tìm α. Tính EX, EY.

b) Tính $\text{cov}(X,Y)$, $\rho(X,Y)$.

c) X và Y có độc lập không.

3.23 Giả sử X là biến ngẫu nhiên rời rạc có phân bố Poisson tham số λ. Tìm hàm khối lượng xác suất với điều kiện $B = \{X \text{ chẵn}\}$.

3.24 Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{XY}(x,y) = \begin{cases} kx & \text{nếu } 0 < y < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

a) Tìm k.

b) Tìm các hàm mật độ xác suất của X và của Y.

c) X và Y có độc lập không?

3.25 Hàm phân bố xác suất của véc tơ ngẫu nhiên (X,Y) có dạng:

$$F_{XY}(x,y)=\begin{cases}1-e^{-x}-e^{-y}+e^{-x-y} & \text{nếu } x>0, y>0\\ 0 & \text{nếu ngược lại}\end{cases}.$$

Tìm hàm mật độ xác suất đồng thời $f_{XY}(x,y)$ và hàm mật độ có điều kiện $f(x|y)$.

3.26 Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{XY}(x,y)=\frac{k}{(1+x^2)(1+y^2)}.$$

a) Tìm k. Tìm hàm phân bố xác suất đồng thời mật độ của X, Y.

b) X và Y có độc lập không?

c) Tính xác suất để véc tơ ngẫu nhiên *(X,Y)* nhận giá trị nằm trong hình chữ nhật với các đỉnh là *A(1,1); B*($\sqrt{3}$,1*); C(1,0)*và $D(\sqrt{3},0)$.

3.27 Chứng minh rằng hàm phân bố điều kiện và hàm mật độ điều kiện của biến ngẫu nhiên X với điều kiện $B=\{a<X\le b\}$ có dạng:

$$F_X(x\,|\,a<X\le b)=\begin{cases}0 & x\le a\\ \dfrac{F_X(x)-F_X(a)}{F_X(b)-F_X(a)} & a<x\le b;\\ 1 & x>b\end{cases}$$

$$f_X(x\,|\,a<X\le b)=\begin{cases}0 & x\le a\\ f_X(x)/\int\limits_a^b f_X(y)dy & a<x\le b\\ 0 & x>b\end{cases}$$

3.28 Giả sử biến ngẫu nhiên $X \sim \mathbf{N}(0;\sigma^2)$. Tìm hàm mật độ điều kiện, hàm phân bố điều kiện và kỳ vọng điều kiện với điều kiện $B = \{X > 0\}$.

3.29 Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{XY}(x,y) = \begin{cases} \dfrac{1}{2x^2 y} & \text{nếu } x \geq 1 \text{ và } \dfrac{1}{x} \leq y \leq x, \\ 0 & \text{nếu ngược lại} \end{cases}$$

a) Tìm hàm mật độ xác suất của X và của Y,

b) Tìm hàm mật độ xác suất có điều kiện $f(x|y)$ và $f(y|x)$.

3.30 Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{XY}(x,y) = \begin{cases} k(x+y) & \text{nếu } 0 < x < 2,\ 0 < y < 2 \\ 0 & \text{nếu ngược lại} \end{cases}$$

a) Tìm k. Tìm hàm mật độ xác suất của X và của Y. X, Y có độc lập không.

b) Tìm hàm mật độ xác suất có điều kiện $f_{X|Y}(x\,|\,y)$ và $f_{Y|X}(y\,|\,x)$.

c) Tính $P\{0 < Y < 1/2 \,|\, X = 1\}$.

3.31 Cho X và Y là hai biến ngẫu nhiên có phân bố chuẩn đồng thời với: $\mathrm{E}X = 35,\ \mathrm{E}Y = 20,\ \mathrm{D}X = 36,\ \mathrm{D}Y = 16$ và $\rho_{X,Y} = 0{,}8$.

Tìm kỳ vọng và phương sai của $2X - 3Y$.

3.32 Cho hai biến ngẫu nhiên X, Y

a) Chứng minh bất đẳng thức Cauchy-Schwarz:

$$\left(\mathrm{E}[XY]\right)^2 \leq \left(\mathrm{E}\,X^2\right)\left(\mathrm{E}\,Y^2\right).$$

b) Chứng minh hệ số tương quan thỏa mãn: $|\rho_{XY}| \leq 1$.

3.33 Cho X là biến ngẫu nhiên với hàm phân bố $F_X(x)$ và hàm mật độ $f_X(x)$. Đặt $Y = aX + B$, $a \neq 0$ và b là hai hằng số thực. Tìm hàm phân bố và hàm mật độ của Y.

Chương 4

LUẬT SỐ LỚN VÀ ĐỊNH LÝ GIỚI HẠN

GIỚI THIỆU

Lý thuyết xác suất nghiên cứu khả năng xuất hiện của các biến cố ngẫu nhiên. Khi tung một đồng xu ta sẽ không biết mặt sấp hay mặt ngửa xuất hiện nhưng nếu tung nhiều lần thì ta thấy rằng số lần mặt sấp và mặt ngửa xuất hiện là xấp xỉ gần bằng nhau. Như vậy khi thực hiện nhiều lần phép thử ngẫu nhiên ta sẽ tìm được quy luật xuất hiện của biến cố ngẫu nhiên, đấy là một nội dung của luật số lớn. Luật số lớn cũng là cơ sở để định nghĩa xác suất của biến cố thông qua tần suất xuất hiện của biến cố đó.

Luật số lớn nghiên cứu sự hội tụ theo xác suất của dãy các biến ngẫu nhiên.

Luật số lớn đầu tiên của James Bernoulli được công bố năm 1713. Về sau, kết quả này được Poisson, Trêbusép, Markov, Liapunốp mở rộng.

Trong chương này ta xét hai định lý về luật số lớn. Định lý Trêbusép là dạng tổng quát của luật số lớn còn định lý Bernoulli là trường hợp đơn giản nhất áp dụng cho dãy các biến ngẫu nhiên độc lập có phân bố Bernoulli $A(p)$ cho ở công thức (2.9).

Để chứng minh định lý Trêbusép ta sử dụng bất đẳng thức Trêbusép.

Định lý giới hạn trung tâm khảo sát sự hội tụ theo phân bố xác suất của dãy các biến ngẫu nhiên. Định lý giới hạn trung tâm áp dụng cho dãy các biến ngẫu nhiên độc lập $X_1, X_2, \ldots$ có cùng phân bố Bernoulli $A(p)$ ta được định lý Moivre –Laplace.

Áp dụng định lý Moivre –Laplace ta có công thức tính gần đúng phân bố nhị thức.

Luật số lớn và định lý giới hạn trung tâm là cơ sở lý thuyết để giải quyết các bài toán ước lượng và kiểm định giả thiết thống kê.

4.1 CÁC DẠNG HỘI TỤ CỦA DÃY CÁC BIẾN NGẪU NHIÊN

4.1.1 Hội tụ theo xác suất

Định nghĩa 4.1: *Xét dãy biến ngẫu nhiên* $\{X_n\}_{n=1}^{\infty}$ *và biến ngẫu nhiên* X *trong cùng một phép thử. Ta nói rằng dãy các biến ngẫu nhiên* $\{X_n\}_{n=1}^{\infty}$ *hội tụ theo xác suất về biến ngẫu nhiên* X, *ký hiệu* $X_n \xrightarrow[n\to\infty]{P} X$, *nếu:*

$$\forall \varepsilon > 0: \quad \lim_{n\to\infty} P\{|X_n - X| > \varepsilon\} = 0 \tag{4.1}$$

Như vậy nếu dãy các biến ngẫu nhiên $X_1, X_2, ...$ hội tụ theo xác suất về biến ngẫu nhiên X thì với n đủ lớn, thực tế ta có thể coi rằng X_n không khác mấy so với X.

4.1.2 Hội tụ theo phân bố

Định nghĩa 4.2: *Dãy các biến ngẫu nhiên* $\{X_n\}_{n=1}^{\infty}$ *được gọi là hội tụ theo phân bố về biến ngẫu nhiên* X *nếu dãy các hàm phân bố xác suất* $\{F_{X_n}(x)\}_{n=1}^{\infty}$ *hội tụ về hàm phân bố xác suất* $F_X(x)$. *Tức là với mọi* $x \in \mathbb{R}$:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \tag{4.2}$$

Trường hợp dãy các biến ngẫu nhiên rời rạc $\{X_n\}_{n=1}^{\infty}$ và biến ngẫu nhiên rời rạc X có cùng tập giá trị $\mathscr{C} = \{c_1, c_2, ...\}$ thì dãy $\{X_n\}_{n=1}^{\infty}$ hội tụ theo phân bố về biến ngẫu nhiên X khi và chỉ khi với mọi $c_k \in \mathscr{C}$:

$$\lim_{n\to\infty} P\{X_n = c_k\} = P\{X = c_k\} \tag{4.3}$$

4.2 LUẬT SỐ LỚN

4.2.1 Bất đẳng thức Markov

Định lý 4.1: Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó với mọi $a > 0$ ta có:

$$P\{Y \geq a\} \leq \frac{EY}{a}. \tag{4.4}$$

Chứng minh:

a) Trường hợp Y rời rạc có tập giá trị $R_Y = \{y_1, y_2, ...\}$.

Đặt: $R_1 = \{y_i \in R_Y, y_i < a\}$; $R_2 = \{y_i \in R_Y, y_i \geq a\}$.

$$EY = \sum_{y_i \in R_Y} y_i P\{Y = y_i\} \ (*)$$

- Trường hợp tổng (*) có hữu hạn các số hạng thì đương nhiên có thể thay đổi thứ tự lấy tổng

- Trường hợp tổng (*) có vô hạn số hạng thì đây là tổng của một chuỗi số dương, do đó có thể thay đổi thứ tự, vì vậy có thể viết lại:

$$EY = \sum_{y_i \in R_Y} y_i P\{Y = y_i\} = \sum_{y_i \in R_1} y_i P\{Y = y_i\} + \sum_{y_i \in R_2} y_i P\{Y = y_i\}$$
$$\geq \sum_{y_i \in R_2} y_i P\{Y = y_i\} \geq a \sum_{y_i \in R_2} P\{Y = y_i\} = aP\{Y \geq a\}.$$

Suy ra: $P\{Y \geq a\} \leq \frac{EY}{a}$.

b) Giả sử Y liên tục có hàm mật độ xác suất $f_Y(x)$. Ta có:

$$EY = \int_0^{+\infty} x f_Y(x) dx$$
$$= \int_0^a x f_Y(x) dx + \int_a^{+\infty} x f_Y(x) dx \geq \int_a^{+\infty} x f_Y(x) dx \geq a \int_a^{+\infty} f_Y(x) dx$$
$$= aP\{Y \geq a\}.$$

Suy ra: $P\{Y \geq a\} \leq \frac{EY}{a}$.

4.2.2 Bất đẳng thức Trêbusép

Định lý 4.2: Giả sử X là biến ngẫu nhiên có kỳ vọng và phương sai hữu hạn, khi đó với mọi $\varepsilon > 0$ ta có:

$$P\{|X - EX| \geq \varepsilon\} \leq \frac{DX}{\varepsilon^2}. \quad (4.5)$$

cũng vậy, $P\{|X - EX| < \varepsilon\} \geq 1 - \frac{DX}{\varepsilon^2}$. (4.6)

Chứng minh: Áp dụng công thức (4.4) cho biến ngẫu nhiên $Y = (X - EX)^2$ và $a = \varepsilon^2$ ta có:

$$P\{|X - EX| \geq \varepsilon\} = P\{Y \geq \varepsilon^2\} \leq \frac{EY}{\varepsilon^2} = \frac{E(X - EX)^2}{\varepsilon^2} = \frac{DX}{\varepsilon^2}.$$

Bất đẳng thức (4.5), (4.6) được gọi là bất đẳng thức Trêbusép.

Bất đẳng thức Trêbusép có nhiều ứng dụng. Trước hết nó cho phép ta đánh giá cận trên hoặc cận dưới xác suất để biến ngẫu nhiên X nhận giá trị sai lệch so với kỳ vọng EX không quá ε. Bất đẳng thức Trêbusép có ý nghĩa to lớn về mặt lý thuyết, nó được sử dụng để chứng minh các định lý của luật số lớn.

Ví dụ 4.1: Một cửa hàng muốn ước lượng nhanh chóng sai số của số vải bán ra trong một tháng của mình. Số vải của mỗi khách hàng được làm tròn bởi số nguyên gần nhất (ví dụ số vải thực bán 195,6*m* thì làm tròn là 196*m*). Ký hiệu X_i là sai số giữa số mét vải thực bán và số mét vải đã làm tròn của khách hàng thứ *i*.

Giải: Các sai số $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có phân bố đều trên đoạn $[-0,5; 0,5]$. Khi đó $EX_i = 0, DX_i = \frac{1}{12}$. Sai số tổng cộng trong cả tháng là $S = X_1 + \cdots + X_n$ (trong đó n là số khách hàng mua hàng trong tháng). Ta có:

$$ES = \sum_{i=1}^{n} EX_i = 0,\ DS = \sum_{i=1}^{n} DX_i = \frac{n}{12}.$$

Theo bất đẳng thức Trêbusép, xác suất để sai số vượt quá ε mét sẽ được đánh giá bởi:

$$P\{|S| \geq \varepsilon\} \leq \frac{DS}{\varepsilon^2} = \frac{n}{12\varepsilon^2}.$$

Giả sử có $n = 10^4$ khách hàng trong tháng. Để xác suất $P\{|S| > \varepsilon\}$ bé hơn 0,01 ta phải có:

$$\frac{n}{12\varepsilon^2} \leq 0,01 \text{ hay } \varepsilon \geq \sqrt{\frac{n}{12 \times 0,01}} = 288,67.$$

Vậy ta có thể kết luận: Với xác suất 0,99 sai số giữa số vải thực bán với số vải đã tính tròn không vượt quá 289m, nếu số khách hàng là 1 vạn.

4.2.3 Luật số lớn Trêbusép

Định lý 4.3: Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập, có các kỳ vọng hữu hạn và tất cả các phương sai bị chặn trên bởi hằng số C ($DX_i \leq C; \forall i = 1,2,...$). Khi đó, với $\forall\ \varepsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{EX_1 + \cdots + EX_n}{n}\right| > \varepsilon\right) = 0. \qquad (4.7)$$

Chứng minh: Xét biến ngẫu nhiên $S_n = \dfrac{X_1 + \cdots + X_n}{n}$. Từ giả thiết độc lập của dãy các biến ngẫu nhiên $X_1, X_2, ...$ ta suy ra:

$$ES_n = \frac{EX_1 + \cdots + EX_n}{n};\ DS_n = \frac{DX_1 + \cdots + DX_n}{n^2} \leq \frac{C}{n}.$$

Áp dụng bất đẳng thức Trêbusép (4.5) cho biến ngẫu nhiên S_n ta có:

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \frac{EX_1 + \cdots + EX_n}{n}\right| > \varepsilon\right) \leq \frac{C}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0.$$

Hệ quả 4.1: Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập có cùng có kỳ vọng μ và tất cả các phương sai bị chặn trên bởi hằng số C ($DX_i \leq C; \forall i = 1,2,...$). Khi đó:

$$\frac{X_1+\cdots+X_n}{n} \xrightarrow[n\to\infty]{P} \mu. \tag{4.8}$$

Hệ quả 4.2: Giả sử $X_1, X_2, \ldots$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Khi đó

$$\frac{X_1+\cdots+X_n}{n} \xrightarrow[n\to\infty]{P} \mu. \tag{4.9}$$

Định lý Trêbusép chứng tỏ rằng trung bình số học của các biến ngẫu nhiên độc lập hội tụ theo xác suất về trung bình số học của kỳ vọng tương ứng của nó. Nói cách khác có sự ổn định của trung bình số học của một số lớn các biến ngẫu nhiên xung quanh trung bình số học của các kỳ vọng của các biến ngẫu nhiên ấy. Như vậy mặc dù từng biến ngẫu nhiên độc lập có thể nhận giá trị khác nhiều so với kỳ vọng của chúng, song trung bình số học của một số lớn các biến ngẫu nhiên lại nhận giá trị gần bằng trung bình số học các kỳ vọng của chúng với xác suất rất lớn. Điều đó cho phép dự đoán giá trị trung bình số học của các biến ngẫu nhiên. Chẳng hạn, gieo một con xúc sắc cân đối. Giả sử X là số chấm xuất hiện ở mặt trên con xúc sắc. Ta có $EX = 3,5$ (xem ví dụ 2.22). Một nhà thống kê đã gieo một con xúc sắc cân đối 1 triệu lần (nhờ sự trợ giúp của máy vi tính) và ghi lại số chấm xuất hiện ở mặt trên con xúc sắc. Số trung bình của 1 triệu lần gieo được tìm thấy là:

$$\frac{x_1+\cdots+x_{10^6}}{10^6} \approx 3,500867 \approx 3,5.$$

Định lý Trêbusép có ứng dụng rộng rãi trong nhiều lĩnh vực, chẳng hạn nó chính là cơ sở cho phương pháp đo lường trong vật lý. Để xác định giá trị của một đại lượng vật lý nào đó người ta thường tiến hành đo n lần độc lập và lấy trung bình số học của các kết quả đo làm giá trị thực của đại lượng cần đo. Thật vậy, giả sử xem kết quả của n lần đo là các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$. Ta thấy rằng các biến ngẫu nhiên này độc lập, có cùng kỳ vọng bằng chính giá trị thực

của đại lượng vật lý (giả sử không có sai số hệ thống), các phương sai của chúng đều bị chặn trên bởi bình phương của độ chính xác của thiết bị đo. Do đó theo định lý Trêbusép ta có thể cho rằng trung bình số học của các kết quả đo sẽ sai lệch rất ít so với giá trị thực của đại lượng vật lý với xác suất gần như bằng một.

Định lý Trêbusép còn là cơ sở cho phương pháp mẫu ứng dụng trong thống kê.

4.2.4 Luật số lớn Bernoulli

Xét phép thử ngẫu nhiên $\mathscr{C}$ và A là một biến cố liên quan đến phép thử $\mathscr{C}$. Thực hiện n lần phép thử $\mathscr{C}$ một cách độc lập và gọi k_n là tần số xuất hiện biến cố A trong n phép thử đó. $f_n = \frac{k_n}{n}$ được gọi là tần suất xuất hiện của A trong n phép thử.

Định lý 4.4: Tần suất f_n hội tụ theo xác suất về xác suất p của biến cố A, nghĩa là với mọi $\varepsilon > 0$

$$\lim_{n\to\infty} P\{|f_n - p| < \varepsilon\} = 1. \tag{4.10}$$

Chứng minh: Xét dãy các biến ngẫu nhiên $X_1, X_2, ..., X_n$ xác định như sau:

$$X_k = \begin{cases} 1 & \text{nếu } A \text{ xảy ra ở phép thử thứ } k \\ 0 & \text{nếu } A \text{ không xảy ra ở phép thử thứ } k \end{cases}$$

Như vậy dãy các biến ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập có cùng phân bố Bernoulli tham số p (công thức (2.17)):

$$\mathrm{E}X_k = p,\ \mathrm{D}X_k = p(1-p).$$

Ta có: $\frac{X_1 + \cdots + X_n}{n} = \frac{k_n}{n} = f_n$.

Vậy theo hệ quả 4.2 của định lý 4.2 suy ra f_n hội tụ theo xác suất về p.

Định lý Bernoulli chỉ ra rằng tần suất xuất hiện của biến cố trong n phép thử độc lập sẽ hội tụ theo xác suất về xác suất của biến cố đó khi số lần thử tăng lên vô hạn. Chính vì vậy định lý Bernoulli là cơ sở lý thuyết của định nghĩa thống kê về xác suất.

Ở thế kỷ 18, nhà toán học Pháp Buffon gieo một đồng tiền 4040 lần và ghi được 2048 lần xuất hiện mặt ngửa, tần suất là 0,507. Một nhà thống kê người Anh gieo đồng tiền 12000 lần và thu được 6019 lần xuất hiện mặt ngửa, tần suất tương ứng 0,5016. Trong một thí nghiệm khác, ông ta gieo 24000 lần và thu được 12012 lần xuất hiện mặt ngửa, tần suất tương ứng là 0,5005. Như vậy ta thấy rằng khi số phép thử tăng lên thì tần suất tương ứng sẽ càng gần 0,5.

4.3 ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Theo công thức (2.61), (2.71) ta có:

$$\mathrm{E}\left[X_1 + \cdots + X_n\right] = n\mu \text{ và } \mathrm{D}\left[X_1 + \cdots + X_n\right] = n\sigma^2$$

Do đó:

$$S_n = \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + \cdots + X_n - \mathrm{E}\left[X_1 + \cdots + X_n\right]}{\sqrt{\mathrm{D}\left[X_1 + \cdots + X_n\right]}}$$

có $\mathrm{E}S_n = 0$ và $\mathrm{D}S_n = 1$.

Định lý 4.5: Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố, có kỳ vọng μ và phương sai σ^2. Khi đó dãy biến ngẫu nhiên $S_n = \dfrac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$ hội tụ theo phân bố về phân bố chuẩn tắc $\mathbf{N}(0;1)$, tức là:

Với $\forall x \in 3$, $\lim\limits_{n\to\infty} F_{S_n}(x) = \Phi(x)$ (4.11)

$\Phi(x)$ là hàm phân bố xác suất của phân bố chuẩn tắc $\mathbf{N}(0;1)$.

Áp dụng định lý giới hạn trung tâm cho dãy các biến ngẫu nhiên độc lập $X_1, X_2, ...$ có cùng phân bố Bernoulli tham số p (công thức (2.9)) ta được định lý Moivre –Laplace:

Định lý 4.6 (Moivre –Laplace): Đối với dãy các biến ngẫu nhiên $X_1, X_2, ...$ độc lập có cùng phân bố Bernoulli tham số p thì:

$$\text{Với mọi } x \in \mathbb{R}, \lim_{n\to\infty} P\left\{\frac{X_1 + \cdots + X_n - np}{\sqrt{npq}} \le x\right\} = \Phi(x). \qquad (4.12)$$

4.4 XẤP XỈ PHÂN BỐ NHỊ THỨC

4.4.1 Xấp xỉ phân bố nhị thức bằng phân bố chuẩn

Giả sử $X_1, X_2, ..., X_n$ độc lập có cùng phân bố Bernoulli tham số p. Theo công thức (2.17) và (2.20) ta có:

$$U_n = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n, p)$$

Công thức (2.18) cho phép tính xác suất:

$$P\{U_n = k\} = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

Tuy nhiên khi n khá lớn ta không thể áp dụng công thức này để tính mà cần đến công thức xấp xỉ (vì sẽ tràn bộ nhớ khi tính toán có sử dụng máy tính).

Định lý 4.7 (*Định lý giới hạn địa phương*): Giả sử X là biến ngẫu nhiên có phân bố nhị thức $\mathcal{B}(n, p)$. Đặt $x_k = \frac{k - np}{\sqrt{npq}}$, khi đó:

$$P_n(k; p) = P\{X = k\} = \varphi\left(\frac{k - np}{\sqrt{npq}}\right)\frac{1}{\sqrt{nqp}}\left(1 + \varepsilon_{n,k}\right) \qquad (4.13)$$

trong đó $\left|\varepsilon_{n,k}\right| < \frac{C}{\sqrt{n}}$ với C là hằng số.

Như vậy khi n đủ lớn ta có thể xấp xỉ:

$$P\{X=k\} \approx \varphi\left(\frac{k-np}{\sqrt{npq}}\right)\frac{1}{\sqrt{nqp}} = \frac{1}{\sqrt{nqp}}\frac{1}{\sqrt{2\pi}}e^{\frac{-(k-np)^2}{2npq}} \qquad (4.14)$$

Để tính xấp xỉ giá trị của hàm phân bố xác suất nhị thức ta có thể áp dụng định lý Moivre-Laplace - công thức (4.12) như sau:

$$P\{U_n \le x\} = P\left\{\frac{U_n-np}{\sqrt{npq}} \le \frac{x-np}{\sqrt{npq}}\right\} \approx \Phi\left(\frac{x-np}{\sqrt{npq}}\right) \qquad (4.15)$$

Người ta thấy rằng xấp xỉ là tốt khi np và nq lớn hơn 5 hoặc khi $npq > 20$.

Công thức (4.15) xấp xỉ phân bố rời rạc bởi phân bố liên tục, vì vậy để xấp xỉ tốt hơn người ta thường sử dụng công thức có dạng sau:

$$P\{U_n \le x\} \approx \Phi\left(\frac{x+\frac{1}{2}-np}{\sqrt{npq}}\right) \qquad (4.16)$$

Ví dụ 4.2: Gieo 3200 lần một đồng xu cân đối và đồng chất. Gọi X là số lần xuất hiện mặt sấp trong 3200 lần gieo đó.

a) Tìm số lần xuất hiện mặt sấp có khả năng nhất. Tính xác suất tương ứng.

b) Tính xác suất $P\{1610 \le X \le 1650\}$.

Giải: a) Ta có: $n = 3200$, $p = 0,5 \Rightarrow (n+1)p = 1600,5$. Vậy số lần xuất hiện mặt sấp có khả năng nhất là 1600 với xác suất tương ứng

$$P_{3200}(1600;0,5) = \frac{3200!}{160011600!}0,5^{3200}$$ (không thể tính được vì tràn bộ nhớ).

Mặt khác nếu tính gần đúng ta có: $x_m = \frac{1600-3200\cdot 0,5}{\sqrt{3200\cdot 0,5\cdot 0,5}} = 0$. Do đó:

$$P_{3200}(1600;0,5) \approx \frac{1}{\sqrt{3200\cdot 0,5\cdot 0,5}}\varphi(0) = \frac{1}{40\sqrt{\pi}} \approx 0,014.$$

b) $P\{10 \le X - 1600 \le 50\}$

$$= P\left\{\frac{10}{20\sqrt{2}} \le \frac{X-1600}{20\sqrt{2}} \le \frac{50}{20\sqrt{2}}\right\} \approx \Phi(1,7678) - \Phi(0,3536)$$

Tra bảng ta có $\Phi(1,7678) - \Phi(0,3536) \approx 0,9616 - 0,6406 = 0,321$.

4.4.2 Xấp xỉ phân bố nhị thức bằng phân bố Poisson

Khi điều kiện xấp xỉ phân bố nhị thức bằng phân bố chuẩn không thỏa mãn (np và nq lớn hơn 5 hoặc khi $npq > 20$), ta có thể xấp xỉ phân bố nhị thức bằng phân bố Poisson.

Định lý 4.8: Cho $X_1, X_2, ...$ là dãy các biến ngẫu nhiên có cùng phân bố nhị thức, ở đó với mỗi n, X_n có phân bố nhị thức $\mathcal{B}(n, p_n)$. Nếu tồn tại giới hạn $\lim\limits_{n\to\infty} np_n = \lambda > 0$ thì X_n hội tụ theo phân bố về biến ngẫu nhiên X có phân bố Poisson tham số λ.

Trong thực tế khi $n > 50$ và $p < 0,1$ người ta có thể xấp xỉ $\mathcal{B}(n; p)$ với phân bố Poisson $\mathcal{P}(\lambda)$ tham số $\lambda = np$:

$$P\{X_n = k\} \approx e^{-np}\frac{(np)^k}{k!}. \qquad (4.17)$$

Ví dụ 4.3: Giả sử xác suất để làm ra mỗi đinh ốc không đúng quy cách là $p = 0,015$. Người ta xếp đinh ốc vào từng hộp, mỗi hộp 100 chiếc.

a) Tính tỉ lệ hộp chứa toàn đinh ốc đúng quy cách.

b) Cần phải xếp ít nhất bao nhiêu đinh ốc trong mỗi hộp để tỉ lệ hộp chứa 100 đinh ốc tốt tối thiểu là 80%.

Giải: a) Nếu gọi X là số đinh ốc không đúng quy cách trong hộp chứa 100 đinh ốc thì $X \sim \mathcal{B}(n; p)$ với $n = 100,\ p = 0,015$.

Áp dụng công thức tính gần đúng (4.17) ta được:

$$P\{X = 0\} \approx e^{-np}\frac{(np)^0}{0!} = e^{-1,5} = 0,2231$$

Vậy có khoảng 22,3% số hộp chứa 100 đinh ốc tốt.

b) Giả sử mỗi hộp chứa $100+k$ đinh ốc, k là số tự nhiên. Gọi X là số đinh ốc không đúng quy cách trong mỗi hộp chứa $100+k$ đinh ốc. $X \sim \mathcal{B}(n;p)$ với $n=100+k$, $p=0,015$. Ta phải xác định k nhỏ nhất để:

$$P\{X \le k\} = \sum_{i=0}^{k} C_n^i (0,015)^i (0,985)^{n-i} \ge 0,8 \cdot$$

Dùng công thức xấp xỉ:

$$P\{X=i\} = C_n^i (0,015)^i (0,985)^{n-i} \approx e^{-\lambda} \frac{\lambda^i}{i!}$$

ở đây $\lambda = np = (100+k)(0,015) \approx 1,5+0,015k \approx 1,5$ (vì k nhỏ).

Vậy cần tìm k nhỏ nhất để

$$e^{-1,5}\left\{1+\frac{1,5}{1!}+\frac{1,5^2}{2!}+\cdots+\frac{1,5^k}{k!}\right\} \ge 0,8$$

$$\Leftrightarrow \quad 1+\frac{1,5}{1!}+\frac{1,5^2}{2!}+\cdots+\frac{1,5^k}{k!} \ge 0,8e^{1,5} = 3,5853.$$

Thử với $k=1,2,...$, ta thấy $k=2$ bất đẳng thức trên được thoả mãn. Như vậy dùng xấp xỉ Poisson ta có thể kết luận mỗi hộp cần đóng 102 chiếc đinh ốc. Khi đó xác suất để có ít nhất 100 đinh ốc tốt trong hộp 102 chiếc là 0,8022.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

4.1 Luật số lớn kết luận về sự hội tụ theo xác suất của trung bình cộng các biến ngẫu nhiên độc lập về trung bình cộng của kỳ vọng của chúng nếu các phương sai của các biến ngẫu nhiên này bị chặn.

Đúng ☐ Sai ☐ .

4.2 Giả sử $\{X_n\}$ là dãy các biến ngẫu nhiên có kỳ vọng bằng nhau và phương sai dần tới 0, khi đó dãy sẽ hội tụ theo xác suất đến kỳ vọng chung của dãy biến ngẫu nhiên trên.

Đúng ☐ Sai ☐ .

4.3 Bất đẳng thức Trêbưsép chỉ đúng đối với các biến ngẫu nhiên rời rạc.

Đúng ☐ Sai ☐ .

4.4 Bất đẳng thức Trêbưsép chỉ đúng đối với các biến ngẫu nhiên nhận giá trị dương.

Đúng ☐ Sai ☐ .

4.5 Luật số lớn Bernoulli là một trường hợp đặc biệt của luật số lớn Trêbưsép khi dãy các biến ngẫu nhiên được có cùng phân bố Bernoulli A(p) .

Đúng ☐ Sai ☐ .

4.6 Luật số lớn Bernoulli là cơ sở lý thuyết của định nghĩa thống kê về xác suất.

Đúng ☐ Sai ☐ .

4.7 Tổng của dãy các biến ngẫu nhiên độc lập có cùng phân bố với kỳ vọng và phương sai hữu hạn tiệm cận phân bố chuẩn.

Đúng ☐ Sai ☐ .

4.8 Luật số lớn xét sự hội tụ theo xác suất còn định lý giới hạn trung tâm xét sự hội tụ theo phân bố của dãy các biến ngẫu nhiên.

Đúng ☐ Sai ☐ .

4.9 Có 10 máy hoạt động độc lập với nhau. Xác suất để trong ca làm việc mỗi máy bị hỏng là 0,05. Dựa vào bất đẳng thức Trêbusép hãy đánh giá xác suất của sự sai lệch giữa số máy hỏng và số máy hỏng trung bình.

a) Nhỏ hơn 2.

b) Lớn hơn 2

4.10 Cho $X_1, X_2, ..., X_{12}$ là các biến ngẫu nhiên độc lập với $\mathrm{E}X_i = 16$, $\mathrm{D}X_i = 1$ $(i = \overline{1,12})$. Sử dụng bất đẳng thức Trêbusép để tìm hai hằng số a, b sao cho:

$$P\left\{a \le \sum_{i=1}^{12} X_i \le b\right\} \ge 0{,}99 .$$

4.11 Cho $X_1, X_2, ..., X_{10000}$ là các biến ngẫu nhiên độc lập có phân bố đều trong đoạn $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Chứng minh rằng:

$$P\left\{\left|\sum_{i=1}^{10000} X_i\right| \ge 500\right\} \le \frac{1}{300}.$$

4.12 Xét biến ngẫu nhiên liên tục X có hàm mật độ $f_X(x)$ và giả sử $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố với X. Ta gọi $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên kích thước n của X. Trung bình mẫu được định nghĩa là $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Ký hiệu $\mathrm{E}X = \mu$, $\mathrm{D}X = \sigma^2$. Cần phải lấy mẫu kích thước bao nhiêu để $P\left\{\left|\overline{X} - \mu\right| \le \frac{\sigma}{10}\right\} \ge 0{,}95$.

4.13 Gieo một con xúc xắc cân đối n lần một cách độc lập. Gọi S là số lần xuất hiện mặt lục. Chứng minh rằng :

$$P\left\{\frac{n}{6}-\sqrt{n}<S<\frac{n}{6}+\sqrt{n}\right\}\geq\frac{31}{36}.$$

4.14 Giả sử tiền điện của một gia đình phải trả trong 1 tháng là một biến ngẫu nhiên với trung bình 16USD và độ lệch tiêu chuẩn 1USD. Sử dụng bất đẳng thức Trêbưsép, hãy xác định số M nhỏ nhất để với xác suất 0,99 số tiền điện phải trả trong 1 năm (12 tháng) không vượt quá M.

4.15 Cho dãy các biến ngẫu nhiên độc lập $\{X_n\}$ xác định như sau:

X_n	-2^n	0	2^n
P	$2^{-(2n+1)}$	$1-2^{-2n}$	$2^{-(2n+1)}$

Dãy $\{X_n\}$ thỏa mãn luật số lớn Trêbưsép không?

4.16 Cho dãy các biến ngẫu nhiên độc lập $\{X_n\}$ xác định như sau:

X_n	$-na$	0	na
P	$\frac{1}{2n^2}$	$1-\frac{1}{n^2}$	$\frac{1}{2n^2}$

trong đó a là một hằng số. Dãy $\{X_n\}$ thỏa mãn luật số lớn Trêbưsép không?

4.17 Cho dãy các biến ngẫu nhiên độc lập $\{X_n\}$ xác định như sau:

X_n	$-a$	a
P	$\frac{n+1}{2n+1}$	$\frac{n}{2n+1}$

trong đó a là một hằng số. Dãy $\{X_n\}$ thỏa mãn luật số lớn Trêbưsép không?

4.18 Xác suất chậm tàu của mỗi hành khách là 0,007. Dùng bất đẳng thức Trêbưsép hãy đánh giá xác suất để trong 20.000 hành khách có từ 100 đến 180 người chậm tàu.

4.19 Phải kiểm tra bao nhiêu chi tiết để với xác suất không nhỏ hơn 0,98 có thể hy vọng rằng sai lệch giữa tần suất xuất hiện chi tiết tốt và xác suất để chi tiết là tốt bằng 0,95 sẽ không vượt quá 0,01.

4.20 Một xí nghiệp sản xuất máy tính có xác suất làm ra sản phẩm phế phẩm là 0,02. Chọn ngẫu nhiên 250 máy tính để kiểm tra. Tính xác suất để:

a) Có đúng hai máy phế phẩm;

b) Có không quá hai máy phế phẩm.

4.21 Một nhà nghỉ có 1000 khách. Nhà ăn phục vụ bữa trưa làm hai đợt liên tiếp. Số chỗ ngồi của nhà ăn phải ít nhất là bao nhiêu để xác suất của biến cố "không đủ chỗ cho người đến ăn" bé hơn 1%?

4.22 Một trường đại học có chỉ tiêu tuyển sinh là 300.

a) Giả sử có 325 người dự thi và xác suất thi đỗ của mỗi người là 90%. Tính xác suất để số người trúng tuyển không vượt quá chỉ tiêu.

b) Cần cho phép tối đa bao nhiêu người dự thi (xác suất đỗ của họ vẫn là 90%) để biến cố "số người trúng tuyển không vượt quá chỉ tiêu tuyển sinh" có xác suất lớn hơn 0,99.

4.23 Một kênh truyền bị nhiễu có xác suất lỗi trên mỗi ký số là $0,01$.

a) Tính xác suất có nhiều hơn một lỗi trong 10 ký số nhận được;

b) Tính kết quả của ý a) bằng cách sử dụng công thức xấp xỉ Poisson.

4.24 Giả sử X là biến ngẫu nhiên có phân bố Poisson tham số λ. Áp dụng định lý giới hạn trung tâm hãy xây dựng công thức xấp xỉ

$$P\{X \le x\} \approx \Phi\left(\frac{x-\lambda}{\sqrt{\lambda}}\right).$$

Chương 5

THỐNG KÊ TOÁN HỌC

GIỚI THIỆU

Thống kê toán là bộ môn toán học nghiên cứu qui luật của các hiện tượng ngẫu nhiên có tính chất số lớn trên cơ sở thu thập và xử lý số liệu thống kê các kết quả quan sát về những hiện tượng ngẫu nhiên này.

Nếu ta thu thập được tất cả các số liệu liên quan đến đối tượng cần nghiên cứu thì ta có thể biết được đối tượng này. Tuy nhiên trong thực tế điều đó không thể thực hiện được vì quy mô của đối tượng nghiên cứu quá lớn hoặc trong quá trình nghiên cứu đối tượng nghiên cứu bị phá hủy. Vì vậy cần lấy mẫu để nghiên cứu.

Phương pháp mẫu là một trong những phương pháp quan trọng của lý thuyết thống kê.

Trong chương này chúng ta tìm hiểu một số vấn đề cơ bản của lý thuyết thống kê toán học:

- Cơ sở của lý thuyết mẫu, các phương pháp chọn mẫu: mẫu ngẫu nhiên đơn, mẫu ngẫu nhiên hệ thống, mẫu chùm, mẫu phân tổ, mẫu nhiều cấp;

- Lý thuyết ước lượng (Estimation theory);

- Kiểm định giả thiết thống kê (Hypothesis testing).

Đối với mẫu ngẫu nhiên ta xét các vấn đề:

- Các phương pháp mô tả mẫu: bảng phân bố tần số và tần suất thực nghiệm, bảng phân bố ghép lớp; biểu đồ tần số hình gậy, đa giác tần suất và tổ chức đồ.

- Thống kê của mẫu ngẫu nhiên.

- Các đặc trưng của thống kê mẫu ngẫu nhiên.

- Quy luật phân bố xác suất của một số thống kê đặc trưng mẫu.

Sử dụng phương pháp quy nạp thống kê ta có thể ước lượng các tham số đặc trưng của tổng thể thông qua thống kê của mẫu.

Phương pháp ước lượng điểm chủ trương dùng giá trị quan sát của một thống kê để ước lượng một tham số nào đó theo các tiêu chuẩn: vững, không chệch, hiệu quả. Có hai phương pháp ước lượng điểm là phương pháp mô-men và phương pháp hợp lý cực đại.

Khoảng tin cậy là khoảng mà tham số của dấu hiệu nghiên cứu của tổng thể rơi vào khoảng này với xác suất bằng độ tin cậy.

Một dạng khác của quy nạp thống kê là kiểm định giả thiết thống kê. Đây là một phương pháp quan trọng cho phép giải quyết nhiều bài toán trong thực tế. Nội dung của kiểm định giả thiết thống kê là dựa vào mẫu cụ thể và các quy tắc hay thủ tục quyết định dẫn đến bác bỏ hay chấp nhận giả thiết của tổng thể.

Trong chương này ta sẽ xây dựng ước lượng cho kỳ vọng, phương sai của dấu hiệu nghiên cứu có phân bố chuẩn và ước lượng cho tần suất của tổng thể. Kiểm định các tham số đặc trưng: kỳ vọng của biến ngẫu nhiên gốc trong tổng thể, tần suất của tổng thể.

Để học tốt chương này bạn đọc cần nắm vững cơ sở lý thuyết xác suất đã được học trong các chương trước.

5.1 LÝ THUYẾT MẪU

5.1.1 Sự cần thiết phải lấy mẫu

Nhiều bài toán trong thực tế dẫn đến nghiên cứu một hay nhiều dấu hiệu định tính hoặc định lượng đặc trưng cho các phần tử của một tập hợp nào đó. Chẳng hạn nếu muốn điều tra thu nhập bình quân của các gia đình ở Hà Nội thì tập hợp cần nghiên cứu là các hộ gia đình ở Hà Nội, dấu hiệu nghiên cứu là thu nhập của từng gia đình. Một doanh nghiệp muốn nghiên cứu các khách hàng của mình về dấu hiệu định

tính có thể là mức độ hài lòng của khách hàng đối với sản phẩm hoặc dịch vụ của doanh nghiệp, còn dấu hiệu định lượng là số lượng sản phẩm của doanh nghiệp mà khách hàng có nhu cầu được đáp ứng. Khi khảo sát một tín hiệu là quá trình ngẫu nhiên người ta tiến hành lấy mẫu tại những thời điểm nào đó và thu được các tín hiệu mẫu.

Để xử lý dấu hiệu cần nghiên cứu đôi khi người ta sử dụng phương pháp nghiên cứu toàn bộ, đó là điều tra toàn bộ các phần tử của tập hợp theo dấu hiệu cần nghiên cứu để rút ra các kết luận cần thiết. Tuy nhiên trong thực tế việc áp dụng phương pháp này gặp phải những khó khăn sau:

- Do qui mô của tập hợp cần nghiên cứu quá lớn nên việc nghiên cứu toàn bộ sẽ đòi hỏi nhiều chi phí về vật chất và thời gian, có thể không kiểm soát được dẫn đến bị chồng chéo hoặc bỏ sót.

- Trong nhiều trường hợp không thể nắm được toàn bộ các phần tử của tập hợp cần nghiên cứu, do đó không thể tiến hành toàn bộ được.

- Có thể trong quá trình điều tra sẽ phá hủy đối tượng nghiên cứu…

Vì thế trong thực tế phương pháp nghiên cứu toàn bộ thường chỉ áp dụng đối với các tập hợp có qui mô nhỏ, còn chủ yếu người ta sử dụng phương pháp không toàn bộ mà đặc biệt là phương pháp nghiên cứu chọn mẫu.

5.1.2 Tổng thể nghiên cứu, dấu hiệu nghiên cứu

Toàn bộ tập hợp các phần tử đồng nhất theo một dấu hiệu nghiên cứu định tính hay định lượng nào đó được gọi là tổng thể, ký hiệu $\mathscr{C}$.

Số lượng các phần tử của tổng thể được gọi là *kích thước của tổng thể*, ký hiệu N. Thường thì kích thước N của tổng thể là hữu hạn, song nếu tổng thể quá lớn hoặc không thể nắm được toàn bộ tổng thể ta có thể giả thiết rằng kích thước của tổng thể là vô hạn.

Mỗi phân tử của tổng thể được gọi là *cá thể*.

Các cá thể của tổng thể được nghiên cứu thông qua các dấu hiệu nghiên cứu. Dấu hiệu nghiên cứu này có thể được định tính hoặc định

lượng. Nếu dấu hiệu nghiên cứu có tính định lượng, nghĩa là được thể hiện bằng cách cho tương ứng mỗi cá thể của tổng thể $\mathscr{C}$ nhận một giá trị thực nào đó thì dấu hiệu này được gọi là một *biến lượng*, ký hiệu X. Bằng cách mô hình hóa ta có thể xem biến lượng X là một biến ngẫu nhiên xác định trên tổng thể $\mathscr{C}$. Đối với dấu hiệu định tính có thể mã hóa và xem là biến ngẫu nhiên có phân bố Bernoulli.

Việc chọn ra từ tổng thể một tập con n cá thể nào đó gọi là *phép lấy mẫu*. Tập hợp con này được gọi là *một mẫu*, n là kích thước mẫu.

5.1.3 Mẫu ngẫu nhiên

Ta nói rằng một mẫu là *mẫu ngẫu nhiên* nếu trong phép lấy mẫu đó mỗi cá thể của tổng thể được chọn một cách độc lập và có xác suất được chọn như nhau.

Giả sử các cá thể của tổng thể được nghiên cứu thông qua dấu hiệu X.

Với mẫu ngẫu nhiên kích thước n, gọi X_i là dấu hiệu X của phần tử thứ i của mẫu ($i=\overline{1,n}$). Bằng cách đồng nhất mẫu ngẫu nhiên với các dấu hiệu nghiên cứu của mẫu ta có định nghĩa về mẫu ngẫu nhiên như sau:

Mẫu ngẫu nhiên kích thước n là một dãy gồm n biến ngẫu nhiên: $X_1, X_2, ..., X_n$ độc lập cùng phân bố với X, ký hiệu:

$$W = (X_1, X_2, ..., X_n).$$

Thực hiện một phép thử đối với mẫu ngẫu nhiên W chính là thực hiện một phép thử đối với mỗi thành phần của mẫu. Giả sử X_i nhận giá trị x_i ($i=\overline{1,n}$), khi đó các giá trị $x_1, x_2, ..., x_n$ tạo thành một giá trị của mẫu ngẫu nhiên, hay còn gọi là một thể hiện của mẫu ngẫu nhiên, ký hiệu $w = (x_1, x_2, ..., x_n)$.

Ví dụ 5.1: Gọi X là số chấm xuất hiện khi tung con xúc xắc cân đối đồng chất, X là biến ngẫu nhiên có bảng phân bố xác suất sau:

X	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

Nếu tung con xúc xắc 3 lần và gọi X_i là số chấm xuất hiện trong lần tung thứ i ($i=\overline{1,3}$) thì ta có 3 biến ngẫu nhiên độc lập có cùng quy luật phân bố xác suất với X. Vậy ta có mẫu ngẫu nhiên kích thước 3, $W=(X_1,X_2,X_3)$.

Thực hiện một phép thử đối với mẫu ngẫu nhiên này tức là tung con xúc xắc 3 lần. Giả sử lần thứ nhất được 2 chấm, lần thứ hai được 5 chấm, lần ba được 3 chấm thì $w=(2,5,3)$ là một mẫu cụ thể của mẫu ngẫu nhiên W.

5.1.4 Các phương pháp mô tả mẫu ngẫu nhiên

5.1.4.1 Bảng phân bố tần số thực nghiệm

Từ một mẫu cụ thể của mẫu ngẫu nhiên kích thước n của X ta sắp xếp các giá trị theo thứ tự tăng dần, giả sử giá trị x_i xuất hiện với tần số r_i, $i=1,...,k$

$$x_1<\cdots<x_k\ ;\ r_1+\cdots+r_k=n. \tag{5.1}$$

Khi đó ta có thể mô tả mẫu ngẫu nhiên trên qua bảng phân bố tần số thực nghiệm:

X	x_1	x_2	...	x_k
Tần số	r_1	r_2	...	r_k

(5.2)

5.1.4.2 Bảng phân bố tần suất thực nghiệm

Ký hiệu $f_i=\dfrac{r_i}{n}$ gọi là tần suất của x_i.

Ta có bảng phân bố tần suất thực nghiệm của X

X	x_1	x_2	...	x_k
Tần suất	f_1	f_2	...	f_k

(5.3)

Ví dụ 5.2: Lấy một mẫu ngẫu nhiên kích thước 120 ta có bảng phân bố tần số thực nghiệm

X	31	34	35	36	38	40	42	44	$\sum$
Tần số	10	20	30	15	10	10	5	20	120

Bảng phân bố tần suất thực nghiệm

X	31	34	35	36	38	40	42	44	$\sum$
Tần suất	2/24	4/24	6/24	3/24	2/24	2/24	1/24	4/24	1

5.1.4.3 Hàm phân bố thực nghiệm của mẫu

Với mẫu mẫu cụ thể của mẫu ngẫu nhiên xác định bởi công thức (5.1), hàm số xác định như sau:

$$F_n(x) = \sum_{x_j \le x} f_j \; ; \quad -\infty < x < +\infty \tag{5.4}$$

gọi là hàm phân bố thực nghiệm của mẫu đã cho.

Định lý Glivenco chỉ ra rằng hàm phân bố thực nghiệm $F_n(x)$ xấp xỉ với phân bố lý thuyết $F_X(x) = P\{X \le x\}$ khi n đủ lớn.

5.1.4.4 Bảng phân bố ghép lớp

Trong những trường hợp mẫu điều tra có kích thước lớn, hoặc khi các giá trị cụ thể của dấu hiệu X lấy giá trị khác nhau song lại khá gần nhau, người ta thường xác định một số các khoảng $C_1, C_2, ..., C_k$ sao cho mỗi giá trị của dấu hiệu điều tra thuộc vào một khoảng nào đó. Các khoảng này lập thành một phân hoạch của miền giá trị của X.

Việc chọn số khoảng và độ rộng khoảng là tuỳ thuộc vào kinh nghiệm của người nghiên cứu, nhưng nói chung không nên chia quá ít khoảng. Ngoài ra độ rộng các khoảng cũng không nhất thiết phải bằng nhau. Chẳng hạn khi muốn thống kê về tỉ lệ người nghiện thuốc lá thì ta tập trung nhiều vào độ tuổi thanh niên và trung niên.

Một trong những gợi ý để chọn số khoảng k tối ưu là hãy chọn k nguyên nhỏ nhất sao cho $2^k \geq n$ như sau:

n : kích thước mẫu	9 –16	17 –32	33 –64	65 –127	129 –256	257 –512
k : số khoảng	4	5	6	7	8	9

Ví dụ 5.3: Một mẫu về chiều cao (cm) của 400 cây con được trình bày trong bảng phân bố ghép lớp sau:

Khoảng	**Tần số r_i**	**Tần suất f_i**	**Độ rộng khoảng l_i**	**$y_i = r_i / l_i$**
4,5-9,5	18	0,,45	5	3,6
9,5-11,5	58	0,145	2	29
11,5-13,5	62	0,155	2	31
13,5-16,5	72	0,180	3	24
16,5-19,5	57	0,1425	3	19
19,5-22,5	42	0,105	3	14
22,5-26,5	36	0,090	4	9
26,5-36,5	55	0,1375	10	5,5

Giá trị $y_i = \dfrac{r_i}{l_i}$ là tần số xuất hiện trong một đơn vị khoảng của khoảng có độ dài l_i.

Qui ước: Đầu mút bên phải của mỗi khoảng thuộc vào khoảng đó mà không thuộc khoảng tiếp theo khi tính tần số của mỗi khoảng.

Trong ví dụ trên ta có các khoảng $[4,5;9,5]$, $(9,5;11,5]$, $(11,5;13,5]$,...

5.1.4.5 Biểu diễn bằng biểu đồ

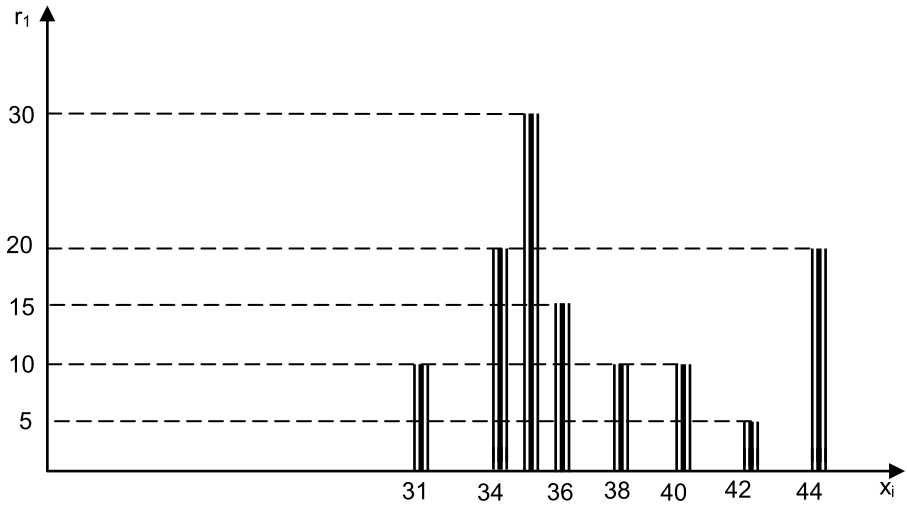
Giả sử dấu hiệu điều tra X có bảng phân bố tần số và tần suất thực nghiệm.

X	x_1	x_2	...	x_k
Tần số	r_1	r_2	...	r_k

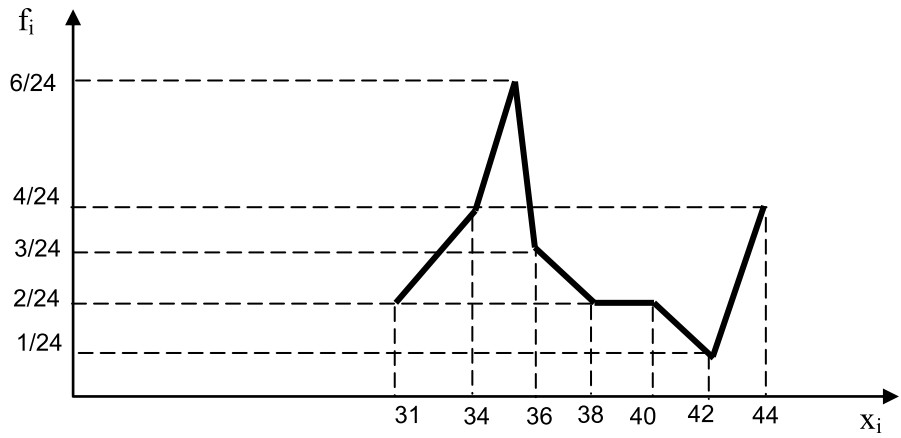
X	x_1	x_2	...	x_k
Tần suất	f_1	f_2	...	f_k

Trong mặt phẳng với hệ trục toạ độ Oxy.

- Nối điểm trên trục hoành có toạ độ $(x_i,0)$ với điểm có toạ độ (x_i,r_i); $i=\overline{1,k}$ ta được *biểu đồ tần số hình gậy*.



Hình 5.1: Biểu đồ tần số hình gậy



Hình 5.2: Biểu đồ đa giác tần suất

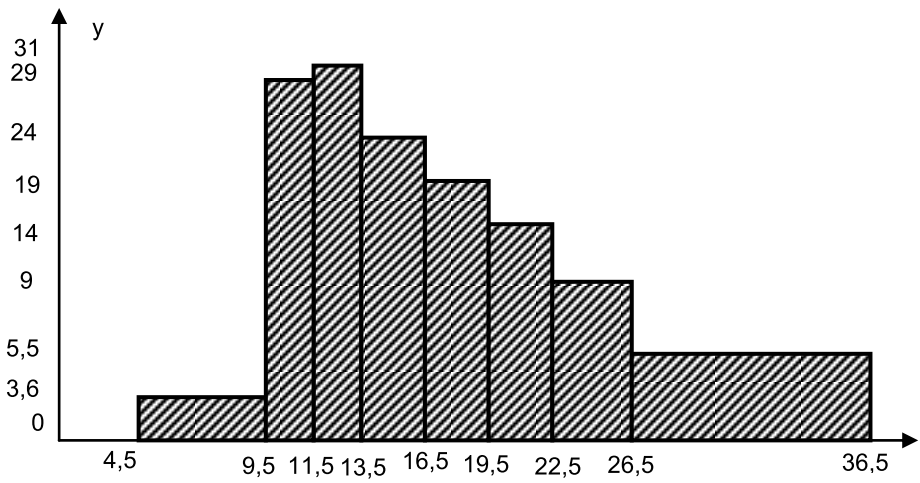
- Nối lần lượt điểm có tọa độ (x_i, f_i) với điểm có tọa độ (x_{i+1}, f_{i+1}); $i = \overline{1, k-1}$ ta được *biểu đồ đa giác tần suất*.

Bảng phân bố tần số và tần suất thực nghiệm trong ví dụ 5.2 có biểu đồ tần số hình gậy và đa giác tần suất hình 5.1 và 5.2.

5.1.4.6 Tổ chức đồ (histogram)

Đối với bảng phân bố ghép lớp, người ta thường dùng tổ chức đồ để biểu diễn.

Trong mặt phẳng với hệ tọa độ Oxy, trên trục hoành ta chia các khoảng C_i có độ rộng l_i. Với mỗi khoảng C_i ta dựng hình chữ nhật có chiều cao $y_i = \dfrac{r_i}{l_i}$ (đối với tổ chức đồ tần số), hay $y_i = \dfrac{f_i}{l_i}$ (đối với tổ chức đồ tần suất). Ví dụ 5.3 có tổ chức đồ hình 5.3.



Hình 5.3: Tổ chức đồ tần số của mẫu ghép lớp của ví dụ 5.3

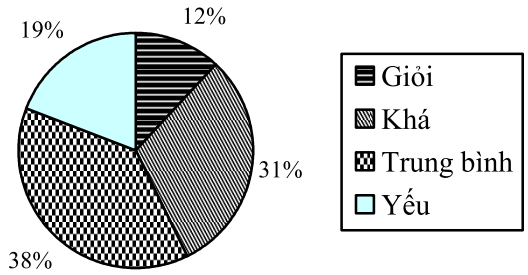
Chú ý rằng diện tích giới hạn bởi tổ chức đồ bằng tần số xuất hiện. Chẳng hạn số cây con nằm trong khoảng (12; 25] chính là diện tích của tổ chức đồ giới hạn bởi đường thẳng $x = 12$ và $x = 25$.

$$(13,5-12)\times 31+(16,5-13,5)\times 24+(19,5-16,5)\times 19$$
$$+(22,5-19,5)\times 14+(25-22,5)\times 9=240$$

Vậy có 240 cây con có chiều cao từ $12cm$ đến $25cm$.

Khi dấu hiệu điều tra có tính chất định tính thì người ta thường mô tả các số liệu mẫu bằng biểu đồ hình bánh xe. Đó là hình tròn được chia thành những góc có diện tích tỷ lệ với các tần số tương ứng của mẫu.

Ví dụ 5.4: Tổng kết kết quả học tập của sinh viên Học viện CNBCVT trong năm 2005 được số liệu sau:



5.1.5 Thống kê và các đặc trưng của mẫu ngẫu nhiên

5.1.5.1 Định nghĩa thống kê

Một thống kê của mẫu là một hàm của các biến ngẫu nhiên thành phần của mẫu. Thống kê của mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ *có dạng:*

$$T=T(X_1,X_2,...,X_n) \tag{5.5}$$

Như vậy thống kê T cũng là một biến ngẫu nhiên tuân theo một quy luật phân bố xác suất nhất định và có các tham số đặc trưng như kỳ vọng ET phương sai DT... (xem mục 3.6). Mặt khác, khi mẫu ngẫu nhiên nhận một giá trị cụ thể $w=(x_1,x_2,...,x_n)$ thì T cũng nhận một giá trị cụ thể còn gọi là giá trị quan sát của thống kê

$$T_{qs}=T(x_1,x_2,...,x_n)$$

Các thống kê cùng với quy luật phân bố xác suất của chúng là cơ sở để suy rộng các thông tin của mẫu cho dấu hiệu nghiên cứu của tổng thể.

5.1.5.2 Trung bình mẫu

Trung bình mẫu của mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ của biến ngẫu nhiên X được định nghĩa và ký hiệu:

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \tag{5.6}$$

Giá trị quan sát của trung bình mẫu của mẫu ngẫu nhiên cụ thể $w=(x_1,x_2,...,x_n)$ là:

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \tag{5.7}$$

Giả sử dấu hiệu nghiên cứu X có kỳ vọng và phương sai hữu hạn, áp dụng các công thức tính kỳ vọng và phương sai của tổng các biến ngẫu nhiên độc lập (2.60), (2.61), (2.71) ta có:

$$\mathrm{E}\left(\overline{X}\right)=\mathrm{E}X;\ \ \mathrm{D}\left(\overline{X}\right)=\frac{\mathrm{D}X}{n}. \tag{5.8}$$

5.1.5.3 Phương sai mẫu

Cho $(X_1,...,X_n)$ là một mẫu ngẫu nhiên của biến ngẫu nhiên gốc X. Ta có các phương sai mẫu ứng với mẫu $(X_1,...,X_n)$ như sau:

- Phương sai mẫu S^2:

$$S^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\frac{1}{n}\sum_{i=1}^{n}{X_i}^2-\left(\overline{X}\right)^2 \tag{5.9a}$$

$$\mathrm{E}\left[S^2\right]=\mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\mu+\mu-\overline{X}\right)^2\right]$$

$$=\frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n}\left(X_i-\mu\right)^2+\left(\overline{X}-\mu\right)^2-2\left(X_i-\mu\right)(\overline{X}-\mu)\right]$$

$$=\frac{1}{n}\mathrm{E}\left[\left(\sum_{i=1}^{n}\left(X_i-\mu\right)^2\right)+n\left(\overline{X}-\mu\right)^2-2n(\overline{X}-\mu)(\overline{X}-\mu)\right]$$

$$= \frac{1}{n}\mathrm{E}\left[\left(\sum_{i=1}^{n}(X_i-\mu)^2\right)-n\left(\overline{X}-\mu\right)^2\right] = \frac{n-1}{n}\mathrm{D}X\,.$$

Để kỳ vọng của phương sai mẫu trùng với kỳ vọng của biến ngẫu nhiên gốc ta cần hiệu chỉnh như sau:

Phương sai mẫu có hiệu chỉnh S^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2\right)-\frac{n}{n-1}\left(\overline{X}\right)^2 \qquad (5.9b)$$

- Phương sai mẫu S^{*2} khi dấu hiệu nghiên cứu X của tổng thể có kỳ vọng xác định $\mathrm{E}X = \mu$.

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \qquad (5.9c)$$

Áp dụng công thức tính kỳ vọng (2.60), (2.61) và (5.8) ta có:

$$\mathrm{E}S^2 = DX \text{ và } \mathrm{E}S^{*2} = \mathrm{D}X \qquad (5.10)$$

5.1.5.4 Độ lệch tiêu chuẩn mẫu

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2} \qquad (5.11)$$

5.1.5.5 Tần suất mẫu

Ta cần nghiên cứu một dấu hiệu định tính A nào đó, mà mỗi cá thể của tổng thể có thể có hoặc không. Nếu cá thể có dấu hiệu A ta cho nhận giá trị 1, trường hợp ngược lại ta cho nhận giá trị 0. Lúc đó dấu hiệu nghiên cứu có thể xem là biến ngẫu nhiên X có phân bố Bernoulli tham số p (xem bảng 2.5.3) có kỳ vọng và phương sai $\mathrm{E}X = p$, $\mathrm{D}X = p(1-p)$ (công thức (2.17)).

Lấy mẫu ngẫu nhiên: $W = (X_1, X_2, ..., X_n)$, $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli với tham số p. Tần số xuất hiện dấu hiệu A của mẫu là:

$$r = X_1 + X_2 + \cdots + X_n\,. \qquad (5.12)$$

Tần suất mẫu

$$f = \frac{r}{n} \tag{5.13}$$

Như vậy tần suất mẫu là trung bình mẫu của biến ngẫu nhiên X có phân bố Bernoulli tham số p.

Tương tự công thức (5.11) ta có công thức tính kỳ vọng và phương sai của tần suất mẫu:

$$\mathrm{E}(f) = p;\ \ \mathrm{D}(f) = \frac{p(1-p)}{n} \tag{5.14}$$

5.1.5.6 Cách tính giá trị cụ thể của trung bình mẫu và phương sai mẫu $\overline{x}$, s^2

1. Nếu mẫu chỉ nhận các giá trị $x_1, x_2, ..., x_k$ với tần số tương ứng $r_1, r_2, ..., r_k$ thì giá trị trung bình mẫu và phương sai mẫu cụ thể được tính theo công thức:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} r_i x_i,\ \ \sum_{i=1}^{k} r_i = n \tag{5.15}$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} r_i \left(x_i - \bar{x}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} r_i x_i^2 - \frac{\left(\sum_{i=1}^{k} r_i x_i\right)^2}{n}\right) \tag{5.16}$$

2. Nếu giá trị của mẫu cụ thể được cho dưới dạng bảng phân bố ghép lớp với các khoảng $C_1, ..., C_m$ và tần số của C_i là r_i thì giá trị trung bình mẫu và phương sai mẫu được tính như trên, trong đó

x_i là trung điểm của khoảng C_i. (5.17)

3. Mẫu thu gọn: nếu các giá trị của mẫu cụ thể x_i không gọn (quá lớn hoặc quá bé hoặc phân tán) ta có thể thu gọn mẫu bằng cách đổi biến:

$$u_i = \frac{x_i - a}{h} \Rightarrow x_i = hu_i + a \ \ \Rightarrow \bar{x} = h\bar{u} + a;\ s_x^2 = h^2 s_u^2 \tag{5.18}$$

Trong đó: $\bar{u} = \frac{1}{n}\sum_{i=1}^{k} r_i u_i$,

$$s_u^2 = \frac{1}{n-1}\sum_{i=1}^{k} r_i\left(u_i - \bar{u}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{k} r_i u_i^2 - \frac{\left(\sum_{i=1}^{k} r_i u_i\right)^2}{n}\right) \quad (5.19)$$

Thật vậy:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} r_i x_i = \frac{1}{n}\sum_{i=1}^{k} r_i\left(hu_i + a\right) = \frac{h}{n}\sum_{i=1}^{k} r_i u_i + \left(\frac{1}{n}\sum_{i=1}^{k} r_i\right)a = h\bar{u} + a .$$

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{k} r_i\left(x_i - \bar{x}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{k} r_i\left(hu_i + a - h\bar{u} - a\right)^2$$

$$= \frac{h^2}{n-1}\sum_{i=1}^{k} r_i\left(u_i - \bar{u}\right)^2 = h^2 s_u^2$$

Các số a và h được chọn phù hợp sao cho $\bar{u}$, s_u^2 tính dễ dàng hơn. Thông thường ta chọn a là điểm giữa của các giá trị x_i.

Ví dụ 5.5: Giá trị trung bình mẫu và phương sai mẫu của mẫu ở ví dụ 5.3.

Khoảng	**Tần số r_i**	x_i	$u_i = \frac{x_i - 20}{5}$	$r_i u_i$	$r_i u_i^2$
4,5-9,5	18	7	-2,6	-46,8	121,68
9,5-11,5	58	10,5	-1,9	-110,2	209,38
11,5-13,5	62	12,5	-1,5	-93	139,5
13,5-16,5	72	15	-1	-72	72
16,5-19,5	57	18	-,04	-22,8	9,12
19,5-22,5	42	21	0,2	8,4	1,68
22,5-26,5	36	24,5	0,9	32,4	29,16
26,5-36,5	55	31,5	2,3	126,5	290,95
$\sum$	400			-177,5	873,47

$$\bar{x} = 5\bar{u} + 20 = 5 \times \frac{-177,5}{400} + 20 = 17,78$$

$$s_u^2 = \frac{1}{399} \times \left(873{,}47 - \frac{(-177{,}5)^2}{400} \right) = 1{,}9917$$

$$s_x^2 = 5^2 \times s_u^2 = 49{,}79 \quad \Rightarrow \quad s = \sqrt{49{,}79} = 7{,}056 .$$

5.1.6 Quy luật phân bố xác suất của một số thống kê đặc trưng mẫu

5.1.6.1 Trường hợp biến ngẫu nhiên gốc có phân bố chuẩn

Giả sử dấu hiệu nghiên cứu trong tổng thể có thể xem như một biến ngẫu nhiên X có phân bố chuẩn với kỳ vọng $EX = \mu$ và phương sai $DX = \sigma^2$. Các tham số này có thể đã biết hoặc chưa biết. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n:

$$W = (X_1, X_2, ..., X_n)$$

Các biến ngẫu nhiên thành phần $X_1, X_2, ..., X_n$ độc lập có cùng phân bố chuẩn $\mathbf{N}(\mu, \sigma^2)$ như X. Theo định lý 3.10 mọi tổ hợp tuyến tính của các biến ngẫu nhiên có phân bố chuẩn là biến ngẫu nhiên có phân bố chuẩn. Vì vậy ta có các kết quả sau:

1) Thống kê trung bình mẫu:

Trung bình mẫu $\overline{X}$ có phân bố chuẩn với kỳ vọng $E(\overline{X}) = \mu$ và phương sai $D(\overline{X}) = \frac{\sigma^2}{n}$. Áp dụng công thức (2.44) suy ra thống kê sau có phân bố chuẩn tắc $\mathbf{N}(0;1)$:

$$U = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \sim \mathbf{N}(0;1) \qquad (5.20)$$

2) Thống kê phương sai S^{*2}.

Từ công thức (5.9c) ta có: $nS^{*2} = \sum_{i=1}^{n} (X_i - \mu)^2$

và $\frac{nS^{*2}}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$.

Vì các biến ngẫu nhiên X_i độc lập nên các biến ngẫu nhiên $\frac{X_i-\mu}{\sigma}$ cũng độc lập. Mặt khác theo (2.44) $\frac{X_i-\mu}{\sigma}\sim \mathbf{N}(0;1)$. Do đó theo công thức (2.51) thống kê $\chi^2=\frac{nS^{*2}}{\sigma^2}$ có phân bố "khi bình phương" n bậc tự do

$$\chi^2=\frac{nS^{*2}}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n) \tag{5.21}$$

3) Thống kê phương sai S^2.

Tương tự thống kê $\frac{(n-1)S^2}{\sigma^2}$ có phân bố "khi bình phương" $n-1$ bậc tự do

$$\chi^2=\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2\sim\chi^2(n-1) \tag{5.22}$$

4) Từ (5.20), (5.22) và U, χ^2 độc lập, áp dụng công thức (2.55) thì $T=\frac{U}{\sqrt{\chi^2/(n-1)}}$ có phân bố Student $n-1$ bậc tự do

$$T=\frac{(\overline{X}-\mu)\sqrt{n}}{S}=\frac{\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}}=\frac{U}{\sqrt{\chi^2/(n-1)}}\sim\mathbf{T}(n-1) \tag{5.23}$$

Quy luật Student $\mathbf{T}(n)$ hội tụ khá nhanh về phân bố chuẩn tắc $\mathbf{N}(0;1)$, do đó trong thực tế nếu $n\geq 30$ ta có thể xem thống kê T xấp xỉ $\mathbf{N}(0;1)$.

5.1.6.2 Quy luật phân bố của tần suất mẫu

Giả sử trong tổng thể dấu hiệu nghiên cứu có thể xem như biến ngẫu nhiên phân bố Bernoulli tham số p. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n:

$$W = (X_1, X_2, ..., X_n)$$

Từ công thức (5.13), (5.14), (5.15) ta đã biết tần suất mẫu $f = \dfrac{X_1 + \cdots + X_n}{n}$ có phân bố nhị thức với các tham số đặc trưng là: $\mathrm{E}(f) = p$; $\mathrm{D}(f) = \dfrac{pq}{n}$.

Áp dụng định lý Moivre-Laplace ta có:

$$\text{Với } \forall x \in 3\,,\ \lim_{n\to\infty} P\left\{\frac{(f-p)\sqrt{n}}{\sqrt{pq}} < x\right\} = \Phi(x) \qquad (5.24)$$

Như vậy có thể xấp xỉ thống kê $U = \dfrac{(f-p)\sqrt{n}}{\sqrt{pq}}$ với phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi n đủ lớn. Người ta thấy rằng xấp xỉ là tốt khi $np > 5$ và $nq > 5$ hoặc $npq > 20$.

$$U = \frac{(f-p)\sqrt{n}}{\sqrt{pq}} \sim \mathbf{N}(0;1) \text{ khi } \begin{cases} np > 5 \\ nq > 5 \end{cases} \text{ hoặc } npq > 20\,. \quad (5.25)$$

5.2 LÝ THUYẾT ƯỚC LƯỢNG

5.2.1 Phương pháp ước lượng điểm

5.2.1.1 Khái niệm ước lượng điểm

Phương pháp ước lượng điểm chủ trương dùng một giá trị để thay cho giá trị của tham số θ chưa biết của tổng thể. Thông thường giá trị được chọn này là giá trị cụ thể của một thống kê Θ nào đó của mẫu ngẫu nhiên.

Với mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$, thống kê ước lượng cho tham số θ được ký hiệu:

$$\Theta = T(X_1, X_2, ..., X_n)$$

Khi đó với mẫu cụ thể $w = (x_1, x_2, ..., x_n)$ giá trị cụ thể của thống kê $\theta = T(x_1, x_2, ..., x_n)$ là ước lượng của θ.

Cùng với một mẫu ngẫu nhiên có thể xây dựng nhiều thống kê Θ khác nhau để ước lượng cho tham số θ. Vì vậy ta cần lựa chọn thống kê tốt nhất để ước lượng cho tham số θ dựa vào các tiêu chuẩn sau.

5.2.1.2 Một số loại ước lượng điểm

5.2.1.2.1 Ước lượng không chệch (unbiased estimator)

Thống kê $\Theta = T(X_1, X_2, ..., X_n)$ là một hàm của các biến ngẫu nhiên $X_1, X_2, ..., X_n$ nên cũng là một biến ngẫu nhiên. Do đó ta có thể xét các đặc trưng của thống kê này.

Định nghĩa 5.1: *Thống kê $\Theta = T(X_1, X_2, ..., X_n)$ được gọi là ước lượng không chệch của θ nếu với mọi giá trị của tham số θ.*

$$\mathrm{E}\left[\Theta(X_1, X_2, ..., X_n)\right] = \theta \qquad (5.26)$$

Nếu $\mathrm{E}(\Theta) \neq \theta$ thì $\Theta = T(X_1, X_2, ..., X_n)$ được gọi là một *ước lượng chệch* của θ.

5.2.1.2.2 Ước lượng hiệu quả (efficient estimator)

Điều kiện (5.26) của ước lượng không chệch có nghĩa rằng trung bình các giá trị của Θ bằng giá trị θ. Tuy nhiên từng giá trị của Θ có thể sai lệch rất lớn so với θ. Vì vậy ta tìm ước lượng không chệch sao cho độ sai lệch trung bình là bé nhất.

Định nghĩa 5.2: *Ước lượng không chệch có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng một mẫu ngẫu nhiên gọi là ước lượng hiệu quả (hay ước lượng phương sai bé nhất).*

Như vậy, để xét xem ước lượng không chệch Θ có phải là ước lượng hiệu quả của θ hay không ta cần phải tìm một cận dưới của phương sai của các ước lượng không chệch và so sánh phương sai của Θ với cận dưới này. Điều này được giải quyết bằng bất đẳng thức Cramer-Rao phát biểu như sau:

Cho mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ được lấy từ tổng thể có dấu hiệu nghiên cứu là biến ngẫu nhiên X mà hàm mật độ xác suất (hay biểu thức xác suất) $f(x,\theta)$ thỏa mãn một số điều kiện nhất định (thường được thỏa mãn trong thực tế, ít ra là các phân bố xác suất đã xét trong chương 2) và Θ là ước lượng không chệch bất kỳ của θ thì:

$$D(\Theta)\geq\frac{1}{nE\left(\frac{\partial(\ln f(X,\theta))}{\partial\theta}\right)^2} \qquad (5.27)$$

Ví dụ 5.6: Dựa vào bất đẳng thức trên ta có thể chứng minh được rằng trung bình mẫu $\overline{X}$ là ước lượng hiệu quả của kỳ vọng μ của dấu hiệu nghiên cứu X của tổng thể có phân bố chuẩn $N(\mu,\sigma^2)$.

Thật vậy theo công thức (5.8) ta có $D(\overline{X})=\frac{\sigma^2}{n}$. Mặt khác theo (2.37) hàm mật độ của X có dạng:

$$f(x,\mu)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\Rightarrow \ln f(x,\mu)=-\ln\left(\sigma\sqrt{2\pi}\right)-\frac{(x-\mu)^2}{2\sigma^2}\Rightarrow\frac{\partial\ln f(x,\mu)}{\partial\mu}=\frac{x-\mu}{\sigma^2}$$

Vậy: $nE\left(\frac{\partial(\ln f(X,\theta))}{\partial\theta}\right)^2=nE\left(\frac{X-\mu}{\sigma^2}\right)^2$

$$=\frac{n}{\sigma^4}E(X-\mu)^2=\frac{n}{\sigma^4}D(X)=\frac{n}{\sigma^2}$$

Như vậy $D(\overline{X})=\frac{\sigma^2}{n}$ đạt giá trị cực tiểu của bất đẳng thức Cramer-Rao, do đó trung bình mẫu $\overline{X}$ là ước lượng hiệu quả của μ.

5.2.1.2.3 Ước lượng vững (consistent estimator)

Định nghĩa 5.3: *Thống kê* $\Theta=T(X_1,X_2,...,X_n)$ *được gọi là ước lượng vững của tham số* θ *của biến ngẫu nhiên gốc* X *nếu* $\Theta=T(X_1,X_2,...,X_n)$ *hội tụ theo xác suất đến* θ *khi* $n\to\infty$.

Nghĩa là với mọi $\varepsilon > 0$, luôn có

$$\lim_{n\to\infty} P\left\{\left|\Theta - \theta\right| < \varepsilon\right\} = 1. \tag{5.28a}$$

Hoặc một cách tương đương

$$\lim_{n\to\infty} P\left\{\left|\Theta - \theta\right| \geq \varepsilon\right\} = 0. \tag{5.28b}$$

Theo hệ quả 4.2 của luật số lớn Trêbưsep, công thức (4.9) chương 4, ta có trung bình mẫu $\overline{X}$ là ước lượng vững của kỳ vọng μ, S^2 và $S^{'2}$ là ước lượng vững của phương sai σ^2 của biến ngẫu nhiên gốc X của tổng thể.

Theo luật số lớn Bernoulli ta suy ra tần suất mẫu f là ước lượng vững của tần suất p của tổng thể.

Tóm lại ta có kết quả sau:

- Trung bình mẫu $\overline{X}$ là ước lượng không chệch, hiệu quả và vững của kỳ vọng μ của biến ngẫu nhiên gốc của tổng thể.

- Tần suất mẫu f là ước lượng không chệch, hiệu quả và vững của tần suất p của tổng thể.

- Phương sai mẫu S^2 và S^{*2} (trường hợp μ đã biết) là ước lượng không chệch và vững của phương sai σ^2 của biến ngẫu nhiên gốc của tổng thể.

5.2.1.3 Một số phương pháp tìm ước lượng điểm

5.2.1.3.1 Phương pháp hợp lý cực đại (maximum-likelihood estimation)

Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên của biến ngẫu nhiên gốc X.

- Nếu X rời rạc với hàm khối lượng xác suất phụ thuộc tham số θ có dạng $p_X(x_k;\theta)$ thì hàm khối lượng xác suất đồng thời của mẫu $(X_1, X_2, ..., X_n)$ xác định như sau:

$$p_{X_1...X_n}(x_{k_1}, ..., x_{k_n};\theta) = p_X(x_{k_1};\theta)...p_X(x_{k_n};\theta) \tag{5.29a}$$

- Nếu X liên tục có hàm mật độ xác suất phụ thuộc tham số θ có dạng $f_X(x;\theta)$ thì hàm mật độ xác suất đồng thời của mẫu $(X_1, X_2, ..., X_n)$:

$$f_{X_1...X_n}(x_1, ..., x_n; \theta) = f_X(x_1; \theta)...f_X(x_n; \theta) \qquad (5.29b)$$

Đặt

$$L(\theta) = \begin{cases} p_{X_1...X_n}(x_{k_1}, ..., x_{k_n}; \theta) & \text{nếu } X \text{ rời rạc} \\ f_{X_1...X_n}(x_1, ..., x_n; \theta) & \text{nếu } X \text{ liên tục} \end{cases} \qquad (5.30)$$

Khi θ là giá trị đúng của tham số thì hàm $L(\theta)$ biểu diễn tính hợp lý về các giá trị $x_1, ..., x_n$ sẽ quan sát được. Vì vậy $L(\theta)$ (θ là biến và $x_1, ..., x_n$ là các tham số) được gọi là hàm hợp lý (likelihood function) của mẫu ngẫu nhiên.

Phương pháp hợp lý cực đại chủ trương coi $\theta = g(x_1, x_2, ..., x_n)$ là ước lượng tốt nhất cho tham số θ, trong đó $\theta = g(x_1, x_2, ..., x_n)$ là điểm cực đại của hàm hợp lý $L(x_1, ..., x_n, \theta)$ theo biên θ và phụ thuộc các tham số $(x_1, x_2, ..., x_n)$.

Thống kê $\theta = g(X_1, X_2, ..., X_n)$ được gọi là ước lượng hợp lý cực đại của θ.

Nói cách khác ước lượng hợp lý cực đại phản ánh nhiều thông tin nhất về tham số θ cần ước lượng.

Giả sử $\theta = g(x_1, ..., x_n)$ là giá trị cực đại của $L(\theta)$, nghĩa là:

$$L(\theta) = \max_{\theta} L(\theta) \qquad (5.31)$$

Thì ước lượng hợp lý cực đại của θ là:

$$\boldsymbol{\Theta} = g(X_1, ..., X_n) \qquad (5.32)$$

Hàm hợp lý $L(\theta)$ là hàm khối lượng xác suất đồng thời hoặc hàm mật độ xác suất đồng thời do đó luôn luôn nhận giá trị dương, ngoài ra hàm lôgarít là hàm đồng biến nên $L(\theta)$ và $\ln L(\theta)$ đạt cực đại tại cùng một giá trị. Vì vậy có thể tìm điểm cực đại θ của hàm $L(\theta)$ bằng cách tìm điểm cực đại của hàm $\ln L(\theta)$.

Hàm hợp lý đạt cực đại tại các điểm dừng θ:

- Nếu $L(\theta)$ là hàm một biến số θ thì θ là nghiệm của phương trình: $\dfrac{d}{d\theta}\ln L(\theta)=0$ (5.33a)

- Nếu $L(\theta)$ là hàm nhiều biến số $\theta=(\theta_1,...,\theta_n)$ thì $\theta=(\theta_1,...,\theta_n)$ là nghiệm của hệ phương trình

$$\begin{cases} \dfrac{\partial}{\partial\theta_1}\ln L(\theta_1,...,\theta_n)=0 \\ ..... \\ \dfrac{\partial}{\partial\theta_n}\ln L(\theta_1,...,\theta_n)=0 \end{cases} \quad (5.33b)$$

Phương pháp này do R. A. Fisher đề xuất và là 1 phương pháp quan trọng để tìm thống kê ước lượng.

Ví dụ 5.7: Giả sử X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, trong đó σ^2 đã biết. Hãy tìm ước lượng hợp lý cực đại của μ.

Trong trường hợp này tham số θ là tham số μ, theo công thức (5.29b), (5.30) ta có:

$$L(\theta)=L(\mu)=\frac{1}{(\sqrt{2\pi}\sigma)^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

Do đó: $\ln L(\theta)=\ln L(\mu)=n\ln\dfrac{1}{\sqrt{2\pi}\sigma}-\dfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$.

$$\frac{d}{d\theta}\ln L(\theta)=\frac{d}{d\mu}\ln L(\mu)=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

Giải phương trình (5.33a) ta được nghiệm $\mu=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$.

Ngoài ra $\frac{d^2}{d\mu^2}\ln L(\mu)=-\frac{n}{\sigma^2}<0$ nên $\ln L(\mu)$ đạt cực đại tại $\mu=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$, nghĩa là thỏa mãn điều kiện (5.31).

Vậy ước lượng hợp lý cực đại của μ là:

$$\boldsymbol{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}.$$

Ví dụ 5.8: Giả sử X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, trong đó μ và σ^2 chưa biết. Hãy tìm ước lượng hợp lý cực đại của μ và σ^2.

Trong trường hợp này tham số θ là $(\mu;\sigma^2)$, theo công thức (5.29b), (5.30) ta có:

$$L(\theta)=L(\mu;\sigma^2)=\frac{1}{(\sqrt{2\pi}\sigma)^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\ln L(\theta)=\ln L(\mu;\sigma^2)=n\ln\frac{1}{\sqrt{2\pi\sigma^2}}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$=-\frac{n}{2}\ln\sigma^2+n\ln\frac{1}{\sqrt{2\pi}}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\begin{cases}\dfrac{\partial}{\partial\mu}\ln L(\mu;\sigma^2)=\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)\\[2ex]\dfrac{\partial}{\partial\sigma^2}\ln L(\mu;\sigma^2)=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2\end{cases}$$

Giải hệ phương trình (5.33b) ta được:

$$\begin{cases} \mu = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \\ \sigma^2 = \dfrac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \hat{s}^2 \end{cases}$$

Xét các đạo hàm riêng cấp 2 tại μ, σ^2:

$$A = \frac{\partial^2}{\partial \mu^2} \ln L(\mu;\sigma^2) = \frac{-n}{\sigma^2};$$

$$C = \frac{\partial^2}{\partial \sigma^4} \ln L(\mu;\sigma^2) = -\frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2 + \frac{n}{2\sigma^4} = -\frac{n}{2\sigma^4}$$

$$B = \frac{\partial^2}{\partial \sigma^2 \partial \mu} \ln L(\mu;\sigma^2) = -\frac{1}{\sigma^4}\sum_{i=1}^{n}(x_i - \mu) = -\frac{1}{\sigma^4}\left\{\left(\sum_{i=1}^{n} x_i\right) - n\mu\right\} = 0.$$

$$A < 0 \text{ và } AC - B^2 = \left(-\frac{n}{\sigma^2}\right)\left(-\frac{n}{2\sigma^4}\right) - 0 = \frac{n^2}{2\sigma^6} > 0.$$

Do đó $L(\theta) = L(\mu;\sigma^2)$ đạt cực đại tại μ, σ^2.

Vậy ta nhận được ước lượng hợp lý cực đại của μ và σ^2 là:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i; \quad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

5.2.1.3.2 Phương pháp mô-men (moment estimation)

Cho mẫu $(X_1, X_2, ..., X_n)$ của biến ngẫu nhiên gốc X, giả sử phân bố của X phụ thuộc vào m tham số $(\theta_1, \theta_2, ..., \theta_m)$ và ta cần ước lượng m tham số này. Tính các mô men của biến ngẫu nhiên gốc ta được hệ

$$\begin{cases} \mathrm{E}\,X = g_1(\theta_1, \theta_2, ..., \theta_m) \\ \mathrm{E}\,X^2 = g_2(\theta_1, \theta_2, ..., \theta_m) \\ ..... \\ \mathrm{E}\,X^m = g_m(\theta_1, \theta_2, ..., \theta_m) \end{cases} \tag{5.34}$$

Thay mô men mẫu vào các vế trái của công thức (5.34) ta được:

$$\begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} X_i = g_1(\theta_1,\theta_2,...,\theta_m) \\ \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 = g_2(\theta_1,\theta_2,...,\theta_m) \\ ..... \\ \dfrac{1}{n}\sum_{i=1}^{n} X_i^m = g_m(\theta_1,\theta_2,...,\theta_m) \end{cases} \tag{5.35}$$

Giải hệ phương trình (5.35) ta được các thống kê $(\Theta_1,\Theta_2,...,\Theta_m)$ của $(X_1,X_2,...,X_n)$.

$(\Theta_1,\Theta_2,...,\Theta_m)$ là ước lượng của $(\theta_1,\theta_2,...,\theta_m)$ theo phương pháp mô men.

Ví dụ 5.9: Giả sử X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, μ,σ^2 đều chưa biết. Hãy tìm ước lượng của μ và σ^2 theo phương pháp mô men.

Hệ phương trình (5.34) có dạng:

$$\begin{cases} \mathrm{E}\,X = \mu \\ \mathrm{E}\,X^2 = \sigma^2 + \mu^2 \end{cases}$$

Suy ra hệ phương trình (5.35)

$$\begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} X_i = \mu \\ \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2 \end{cases}$$

Suy ra: $\mu = \overline{X}$;

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mu^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S^2$$

Trong trường hợp này ước lượng tìm bằng phương pháp mô men trùng với ước lượng tìm bằng phương pháp hợp lý cực đại.

5.2.2 Phương pháp ước lượng bằng khoảng tin cậy

Phương pháp ước lượng điểm nói trên có nhược điểm là khi kích thước mẫu bé thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng. Mặt khác phương pháp trên cũng không thể đánh giá được khả năng mắc sai lầm khi ước lượng là bao nhiêu. Do đó khi kích thước mẫu bé người ta thường dùng phương pháp ước lượng khoảng tin cậy. Theo phương pháp này từ mẫu ngẫu nhiên ta có thể tìm được khoảng [a; b] chứa tham số θ với xác suất β đủ lớn cho trước (β được gọi là độ tin cậy và thường được chọn trong khoảng 0,95 đến 0,99).

Định nghĩa 5.4: *Khoảng* $[a;b]$ *có hai đầu mút là hai thống kê*

$$a = a(X_1, X_2, ..., X_n),\ b = b(X_1, X_2, ..., X_n) \qquad (5.36)$$

phụ thuộc mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$ *của biến ngẫu nhiên gốc* X, *gọi là khoảng tin cậy của tham số* θ *với độ tin cậy* β *nếu:*

$$P\{a(X_1, X_2, ..., X_n) \le \theta \le b(X_1, X_2, ..., X_n)\} = \beta \qquad (5.37)$$

Trong thực tế thường yêu cầu độ tin cậy β khá lớn, khi đó theo nguyên lý xác suất lớn biến cố $\{a \le \theta \le b\}$ hầu như chắc chắn sẽ xảy ra trong một phép thử.

Tiến hành một phép thử với mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$ ta thu được một mẫu cụ thể $w = (x_1, x_2, ..., x_n)$, tính được giá trị cụ thể $a = a(x_1, x_2, ..., x_n)$, $b = b(x_1, x_2, ..., x_n)$.

Lúc đó có thể kết luận là: Qua mẫu cụ thể với độ tin cậy β tham số θ của biến ngẫu nhiên gốc X sẽ nằm trong khoảng $[a;b]$, tức là $a \le \theta \le b$.

5.2.3 Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên phân bố chuẩn

Giả sử biến ngẫu nhiên gốc X có phân bố chuẩn $N(\mu;\sigma^2)$ nhưng chưa biết tham số μ của nó.

Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n: $W=(X_1,X_2,...,X_n)$, ta tìm khoảng tin cậy của μ trong các trường hợp sau:

5.2.3.1 Trường hợp phương sai σ^2 đã biết

Định lý 5.1: Khoảng tin cậy của tham số μ với độ tin cậy β có dạng:

$$\left[\overline{X}-U_{\alpha/2}\frac{\sigma}{\sqrt{n}};\ \overline{X}+U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{5.38}$$

trong đó: $\alpha=1-\beta$; $U_{\frac{\alpha}{2}}$ là giá trị tới hạn mức $\frac{\alpha}{2}$ của phân bố chuẩn tắc $\mathbf{N}(0;1)$ (công thức 2.42).

Chứng minh: Theo công thức (5.20) ta có $\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\sim\mathbf{N}(0;1)$.

Mặt khác:

$$\overline{X}-U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ \Leftrightarrow\ \left|\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\right|\le U_{\alpha/2}.$$

Áp dụng công thức (2.43) ta có

$$P\left[\overline{X}-U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$
$$=P\left\{\left|\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\right|\le U_{\alpha/2}\right\}=1-\alpha=\beta.$$

Định nghĩa 5.5: $\varepsilon=U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ được gọi là *độ chính xác của ước lượng*.

Với độ chính xác ε_0 và độ tin cậy β cho trước, *kích thước mẫu cần thiết* là số tự nhiên n nhỏ nhất thỏa mãn:

$$n\ge\frac{\sigma^2U_{\alpha/2}^2}{\varepsilon_0^2} \tag{5.39}$$

Ví dụ 5.10: Trọng lượng của một loại sản phẩm là một biến ngẫu nhiên có phân bố chuẩn với độ lệch tiêu chuẩn 1 gram. Cần thử 25 sản phẩm loại này ta thu được kết quả:

Trọng lượng (gram)	18	19	20	21
Số sản phẩm tương ứng	3	5	15	2

Với độ tin cậy 95%.

a) Hãy tìm khoảng tin cậy của trọng lượng trung bình của loại sản phẩm trên.

b) Nếu muốn độ chính xác của ước lượng không vượt quá 0,3gram thì cần cân thử ít nhất bao nhiêu sản phẩm.

Giải: Gọi X là trọng lượng sản phẩm, theo giả thiết X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ với $\sigma = 1$. Trọng lượng trung bình của sản phẩm là tham số μ. Khoảng tin cậy có dạng (5.38).

Với độ tin cậy $\beta = 0,95 \Rightarrow \dfrac{\alpha}{2} = 0,025 \Rightarrow U_{\frac{\alpha}{2}} = 1,96$.

a) Từ bảng số liệu tìm được trung bình mẫu cụ thể:

$$\overline{x} = \frac{3\cdot 18 + 5\cdot 19 + 15\cdot 20 + 2\cdot 21}{25} = 19,64 .$$

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} = 1,96\cdot\dfrac{1}{\sqrt{25}} = 0,392$.

Vậy với độ tin cậy 95% qua mẫu cụ thể này, khoảng tin cậy của tham số μ là:

$$[19,64 - 0,392;\ 19,64 + 0,392]$$

hay $19,248 \le \mu \le 20,032$.

b) Nếu muốn độ chính xác của ước lượng không vượt quá 0,3gram thì cần cân thử ít nhất n sản phẩm sao cho:

$$n \ge \frac{\sigma^2 U_{\alpha/2}^2}{\varepsilon_0^2} = \frac{1\cdot 1,96^2}{0,3^2} = 42,68 .$$ Chọn $n = 43$.

5.2.3.2 Trường hợp phương sai σ^2 chưa biết, kích thước mẫu $n \geq 30$

Trong nhiều bài toán thực tế, ta không biết phương sai σ^2 của biến ngẫu nhiên gốc X của tổng thể. Nhưng nếu kích thước mẫu n đủ lớn ($n \geq 30$) ta có thể xấp xỉ độ lệch chuẩn σ bởi độ lệch chuẩn mẫu S (vì S^2 là ước lượng vững không chệch của σ^2), S được xác định bởi công thức (5.11). Mặt khác, theo định lý giới hạn trung tâm thì thống kê $\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}$ có phân bố xấp xỉ phân bố chuẩn, đúng với mọi biến ngẫu nhiên gốc X (không đòi hỏi phân bố chuẩn).

Do đó khoảng tin cậy của tham số μ với độ tin cậy β có thể lấy là:

$$\left[\overline{X} - U_{\alpha/2}\frac{S}{\sqrt{n}};\ \overline{X} + U_{\alpha/2}\frac{S}{\sqrt{n}}\right] \qquad (5.40)$$

Ví dụ 5.11: Để xác định chiều cao trung bình của các cây bạch đàn trong khu rừng trồng bạch đàn, ta tiến hành đo ngẫu nhiên 35 cây và có kết quả cho trong bảng sau:

Khoảng	r_i	x_i	$u_i = x_i - 8,25$	$r_i u_i$	$r_i u_i^2$
6,5-7,0	2	6,75	-1,5	-3	4,5
7,0-7,5	4	7,25	-1,0	-4	4
7,5-8,0	10	7,75	-0,5	-5	2,5
8,0-8,5	11	8,25	0	0	0
8,5-9,0	5	8,75	0,5	2,5	1,25
9,0-9,5	3	9,25	1,0	3	3
$\sum$	35			-6,5	15,25

$$\overline{u} = \frac{-6,5}{35} = -0,1857 \ \Rightarrow\ \overline{x} = 8,25 - 0,1857 \approx 8,06$$

$$s_X^2 = s_U^2 = \frac{1}{34}\left(15,25 - \frac{(-6,5)^2}{35}\right) = 0,413 \ \Rightarrow\ s = 0,64$$

Với độ tin cậy: $\beta = 95\%$, $U_{\alpha/2} = 1,96$.

Độ chính xác của ước lượng: $\varepsilon = U_{\alpha/2}\dfrac{s}{\sqrt{n}} = 1,96 \cdot \dfrac{0,64}{\sqrt{35}} = 0,21$.

Vậy khoảng tin cậy cho chiều cao trung bình μ của các cây bạch đàn là: $7,85 \le \mu \le 8,27$.

5.2.3.3 Trường hợp phương sai σ^2 chưa biết, kích thước mẫu $n < 30$

Trong trường hợp này, theo công thức (5.23) thống kê:

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{S} \tag{5.41}$$

có phân bố Student $n-1$ bậc tự do. Vì vậy khoảng tin cậy được tính theo kết quả sau:

Định lý 5.2: Khoảng tin cậy của tham số μ với độ tin cậy β có dạng:

$$\left[\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}};\ \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right] \tag{5.42}$$

trong đó $t_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\dfrac{\alpha}{2}$ của phân bố Student $n-1$ bậc tự do (công thức 2.56).

Độ chính xác của ước lượng:

$$\varepsilon = t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} \tag{5.43}$$

Với độ tin cậy β và độ chính xác ε_0 cho trước thì kích thước mẫu cần thiết là số tự nhiên n nhỏ nhất thỏa mãn:

$$n \ge \left(\frac{S \cdot t_{\alpha/2}(n-1)}{\varepsilon_0}\right)^2 \tag{5.44}$$

Ví dụ 5.12: Năng suất của một loại giống mới là một biến ngẫu nhiên có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$. Gieo thử giống này trên 16 mảnh vườn thí nghiệm thu được như sau (đơn vị kg/ha):

172, 173, 173,174, 174, 175, 176, 166,
166, 167, 165, 173, 171, 170, 171, 170.

Hãy tìm khoảng tin cậy cho năng suất trung bình của loại giống này với độ tin cậy $\beta = 95\%$.

Giải: Năng suất trung bình của hạt giống là tham số μ.

Từ các số liệu trên ta tính được:

$$\overline{x} = 171;\ s = 3,4254.\ \alpha = 0,05;\ \frac{\alpha}{2} = 0,025.$$

Tra bảng phân bố Student với 15 bậc tự do ta tìm được $t_{\alpha/2}(n-1) = t_{0,025}(15) = 2,131$.

Độ chính xác: $\varepsilon = t_{\alpha/2}(n-1)\dfrac{S}{\sqrt{n}} = 2,131 \cdot \dfrac{3,4254}{\sqrt{16}} = 1,885$.

Vậy khoảng tin cậy cho năng suất trung bình của loại hạt giống này là μ thỏa mãn: $169,115 \le \mu \le 172,885$.

5.2.4 Khoảng tin cậy cho tần suất *p*

Ta cần nghiên cứu một dấu hiệu định tính A nào đó mà mỗi cá thể của tổng thể có thể có hoặc không. Nếu cá thể có dấu hiệu A ta cho nhận giá trị 1, trường hợp ngược lại ta cho nhận giá trị 0. Lúc đó dấu hiệu nghiên cứu có thể xem là biến ngẫu nhiên X có phân bố Bernoulli với tham số p. Kỳ vọng $EX = p$ và phương sai $DX = p(1-p)$.

Lấy mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$, trong đó $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli tham số p.

Tần suất mẫu: $f = \dfrac{X_1 + \cdots + X_n}{n}$.

Theo định lý Moivre-Laplace (4.12) ta có thể xấp xỉ thống kê $U = \dfrac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$ với phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi n đủ lớn.

Tuy nhiên vì p chưa biết nên chưa biết $p(1-p)=\mathrm{D}X$.

Mặt khác theo công thức (5.22), (5.25) thì tần suất mẫu f là ước lượng vững, không chệch và hiệu quả của tần suất p tổng thể. Vì vậy khi n đủ lớn ta có thể thay p bằng f.

Do đó khoảng tin cậy cho tần suất p của tổng thể với độ tin cậy β là:

$$\left[f - U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \; ; \; f + U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \right] \tag{5.45}$$

Với điều kiện: $\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases}$ (5.46)

trong đó $U_{\alpha/2}$ là giá trị tới hạn mức $\frac{\alpha}{2}$ của phân bố chuẩn tắc $\mathbf{N}(0;1)$ với $\alpha = 1-\beta$.

Độ chính xác của khoảng tin cậy:

$$\varepsilon = U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}. \tag{5.47}$$

Với độ tin cậy β và độ chính xác ε_0 cho trước thì *kích thước mẫu cần thiết* là số tự nhiên n nhỏ nhất thỏa mãn:

$$n \geq f(1-f)\left(\frac{U_{\alpha/2}}{\varepsilon_0}\right)^2 \tag{5.48}$$

trong đó f là tần suất mẫu của một mẫu ngẫu nhiên nào đó.

Ví dụ 5.13: Trong đợt vận động bầu cử tổng thống ở một nước nọ, người ta phỏng vấn ngẫu nhiên 1600 cử tri, được biết có 960 người trong số đó sẽ bỏ phiếu cho ứng cử viên A. Với độ tin cậy $\beta = 95\%$ thì ứng cử viên A sẽ chiếm được tối thiểu bao nhiêu % số phiếu bầu.

Giải: Gọi p là tỉ lệ số phiếu sẽ bầu cho ứng cử viên A. Tổng thể nghiên cứu là tập hợp tất cả các cử tri. Dấu hiệu nghiên cứu là cử tri sẽ

bỏ phiếu cho A, là biến ngẫu nhiên có phân bố Bernoulli tham số p. Khoảng tin cậy cho p có dạng (5.45) với điều kiện (5.46).

Từ mẫu cụ thể trên ta có $f = \frac{960}{1600} = 0,6$ thỏa mãn điều kiện:

$$\begin{cases} nf = 960 > 10 \\ n(1-f) = 640 > 10 \end{cases}.$$

Độ chính xác của ước lượng:

$$\varepsilon = U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} = 1,96\sqrt{\frac{0,6 \cdot 0,4}{1600}} = 0,024.$$

Khoảng tin cậy: $0,576 \le p \le 0,624$.

Vậy với độ tin cậy 95% thì tối thiểu có 57,6% cử tri sẽ bỏ phiếu cho ứng cử viên A.

5.2.5 Ước lượng phương sai của biến ngẫu nhiên có phân bố theo quy luật chuẩn $N(\mu;\sigma^2)$

Giả sử trong tổng thể biến ngẫu nhiên gốc X có phân bố chuẩn $N(\mu;\sigma^2)$ nhưng chưa biết phương sai σ^2 của nó. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$.

Ta sẽ chọn thống kê thích hợp để ước lượng cho tham số σ^2 trong các trường hợp sau:

5.2.5.1 Đã biết kỳ vọng μ

Chọn thống kê

$$T = \frac{nS^2}{\sigma^2} \qquad (5.49)$$

Theo công thức (5.21) thống kê T có phân bố "khi bình phương" n bậc tự do: $\chi^2(n)$. Do đó với độ tin cậy β cho trước, với cặp số α_1,

α_2 sao cho $\alpha_1 + \alpha_2 = 1 - \beta$, có thể tìm hai giá trị tới hạn của T mức α_1, α_2 là $\chi^2_{1-\alpha_1}(n)$, $\chi^2_{\alpha_2}(n)$:

$$P\left\{T > \chi^2_{1-\alpha_1}(n)\right\} = 1 - \alpha_1 \quad \text{và} \quad P\left\{T > \chi^2_{\alpha_2}(n)\right\} = \alpha_2 .$$

Do đó:

$$P\left\{\chi^2_{1-\alpha_1}(n) < T < \chi^2_{\alpha_2}(n)\right\} = 1 - (\alpha_1 + \alpha_2) = \beta . \qquad (5.50)$$

Thay thống kê T từ công thức (5.49) vào (5.50) và giải theo σ^2, ta được:

$$P\left\{\frac{nS^2}{\chi^2_{\alpha_2}(n)} < \sigma^2 < \frac{nS^2}{\chi^2_{1-\alpha_1}(n)}\right\} = \beta . \qquad (5.51)$$

Như vậy, với độ tin cậy β khoảng tin cậy của σ^2 có dạng:

$$\left(\frac{nS^2}{\chi^2_{\alpha_2}(n)} ; \frac{nS^2}{\chi^2_{1-\alpha_1}(n)}\right) \qquad (5.52)$$

- Nếu $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ khoảng tin cậy có dạng:

$$\left(\frac{nS^2}{\chi^2_{\alpha/2}(n)} ; \frac{nS^2}{\chi^2_{1-\alpha/2}(n)}\right) \qquad (5.53)$$

- Nếu $\alpha_1 = 0;\ \alpha_2 = \alpha$ khoảng tin cậy bên phải của σ^2 có dạng:

$$\left(\frac{nS^2}{\chi^2_{\alpha}(n)} ; +\infty\right) \qquad (5.54)$$

- Nếu $\alpha_2 = 0; \alpha_1 = \alpha$ khoảng tin cậy bên trái của σ^2 có dạng:

$$\left(0 ; \frac{nS^2}{\chi^2_{1-\alpha}(n)}\right) \qquad (5.55)$$

Ví dụ 5.9: Mức hao phí nguyên liệu cho 1 đơn vị sản phẩm là biến ngẫu nhiên có phân bố chuẩn với trung bình là 20 gam. Để ước lượng mức độ phân tán của mức hao phí này người ta cân thử 25 sản phẩm và thu được kết quả sau:

Hao phí nguyên liệu (gam)	19,5	20,0	20,5
Số sản phẩm tương ứng	5	18	2

Với độ tin cậy $\beta = 90\%$ hãy tìm khoảng tin cậy của σ^2 nếu $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2} = 0,05$.

Giải: Gọi X là mức hao phí nguyên liệu cho 1 đơn vị sản phẩm. X có phân bố chuẩn với kỳ vọng đã biết $\mu = 20$. Đây là ước lượng phương sai σ^2 của phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ khi đã biết μ. Khoảng tin cậy theo công thức (5.53).

Tra bảng $\chi^2(n)$ ta có:

$$\chi^2_{\alpha/2}(n) = \chi^2_{0,05}(25) = 37,65;\ \chi^2_{1-\alpha/2}(n) = \chi^2_{0,95}(25) = 14,61$$

Để tìm $\hat{s}^2$ ta lập bảng sau:

x_i	r_i	$x_i - \mu$	$(x_i - \mu)^2$	$r_i (x_i - \mu)^2$
19,5	5	-0,5	0,25	1,25
20,0	18	0,0	0,00	0,00
20,5	2	0,5	0,25	0,50
$\sum$	25			1,75

$$\hat{s}^2 = \frac{1,75}{25} = 0,07$$

Vậy với độ tin cậy 90%, qua mẫu cụ thể này, khoảng tin cậy của σ^2 là:

$$\left(\frac{25 \cdot 0,07}{37,65}; \frac{25 \cdot 0,07}{14,61}\right) \text{ hay } \quad 0,0464 < \sigma^2 < 0,1198$$

5.2.5.2 Chưa biết kỳ vọng μ

Chọn thống kê

$$T = \frac{(n-1)S^2}{\sigma^2} \tag{5.56}$$

Theo công thức (5.22) thống kê T có phân bố khi bình phương $n-1$ bậc tự do: $\chi^2(n-1)$. Do đó với độ tin cậy β cho trước, với cặp số α_1, α_2 sao cho $\alpha_1+\alpha_2=1-\beta$ có thể tìm hai giá trị tới hạn của T mức α_1, α_2 là $\chi^2_{1-\alpha_1}(n-1), \chi^2_{\alpha_2}(n-1)$:

$$P\left\{T > \chi^2_{1-\alpha_1}(n-1)\right\} = 1-\alpha_1 \quad \text{và} \quad P\left\{T > \chi^2_{\alpha_2}(n-1)\right\} = \alpha_2 .$$

Do đó: $P\left\{\chi^2_{1-\alpha_1}(n-1) < T < \chi^2_{\alpha_2}(n-1)\right\} = 1-(\alpha_1+\alpha_2) = \beta$. (5.57)

Thay thống kê T từ công thức (5.56) vào (5.57) và giải theo σ^2, ta được:

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right\} = \beta . \tag{5.58}$$

Như vậy, với độ tin cậy β khoảng tin cậy của σ^2 có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)} ; \frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right) \tag{5.59}$$

- Nếu $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ khoảng tin cậy có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} ; \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right) \tag{5.60}$$

- Nếu $\alpha_1 = 0; \alpha_2 = \alpha$ khoảng tin cậy bên phải của σ^2 có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)} ; +\infty\right) \tag{5.61}$$

- Nếu $\alpha_2 = 0; \alpha_1 = \alpha$ khoảng tin cậy bên trái của σ^2 có dạng:

$$\left(0 ; \frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}\right) \tag{5.62}$$

5.3 KIỂM ĐỊNH GIẢ THIẾT THỐNG KÊ

5.3.1 Khái niệm chung

Trong mục trước ta giải quyết các bài toán về ước lượng tham số đặc trưng của dấu hiệu nghiên cứu của tổng thể bằng cách đưa về ước lượng các tham số đặc trưng của các biến ngẫu nhiên gốc. Trong mục này ta sẽ nghiên cứu bài toán kiểm định giả thiết về các tham số đặc trưng của tổng thể.

Trước hết ta tìm hiểu các khái niệm sau.

5.3.1.1 Giả thiết thống kê

Giả thiết thống kê là giả thiết về dạng phân bố xác suất, các đặc trưng tham số của biến ngẫu nhiên gốc hoặc giả thiết về sự độc lập của các biến ngẫu nhiên gốc.

Giả thiết đưa ra kiểm nghiệm được ký hiệu là H_0, gọi là *"giả thiết không"* (Null hypothesis). Đó là giả thiết mà ta nghi ngờ muốn bác bỏ hoặc giả thiết ta muốn bảo vệ. Ngoài giả thiết H_0 ra, ta còn phải định ra một giả thiết cạnh tranh với H_0 gọi là *đối thiết* (Alternative hypothesis), ký hiệu H_1. Đối thiết H_1 sẽ được chấp nhận khi H_0 bị bác bỏ.

Cần chú ý rằng đối thiết H_1 không nhất thiết là phủ định của giả thiết H_0. Chẳng hạn giả thiết H_0: nhu cầu thị trường về loại hàng hóa này là $\mu = 1000$ đơn vị/tháng. Nếu ta nghi ngờ rằng nhu cầu này không đúng thì đối thiết H_1 là $\mu \neq 1000$, nhưng nếu do tiếp thị tốt, do chính sách hậu mãi tốt người ta nghĩ rằng nhu cầu về mặt hàng này tăng lên thì đối thiết H_1 là $\mu > 1000$.

Qui tắc kiểm định dựa trên hai nguyên lý sau:

- Nguyên lý xác suất nhỏ: "Nếu một biến cố có xác rất nhỏ thì trong một phép thử ta có thể cho rằng biến cố đó không xảy ra".

- Phương pháp phản chứng: "Ta sẽ bác bỏ A, nếu từ giả thiết A đúng đưa đến một điều vô lý".

Dựa vào hai nguyên lý này ta đưa ra phương pháp chung để kiểm định một giả thiết thống kê như sau: Để kiểm định H_0 trước hết giả sử H_0 đúng từ đó ta tìm được biến cố A mà xác suất xuất hiện biến cố A là rất bé và ta có thể xem A không thể xảy ra trong một phép thử về biến cố này. Lúc đó nếu trên một mẫu cụ thể quan sát được mà biến cố A xuất hiện thì điều này trái với nguyên lý xác suất nhỏ. Vậy H_0 sai và bác bỏ nó. Còn nếu A không xảy ra thì ta chưa có cơ sở để bác bỏ H_0.

Ta thực hiện phương pháp trên bằng các bước cụ thể sau:

5.3.1.2 Tiêu chuẩn kiểm định giả thiết thống kê

Từ biến ngẫu nhiên gốc X của tổng thể lập mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$. Chọn thống kê

$$T = T(X_1, X_2, ..., X_n, \theta) \tag{5.63}$$

trong đó θ là tham số liên quan đến giả thiết cần kiểm định. Nếu H_0 đúng thì thống kê T có quy luật phân bố xác suất xác định. Thống kê T được gọi là *tiêu chuẩn kiểm định.*

5.3.1.3 Miền bác bỏ giả thiết

Sau khi đã chọn tiêu chuẩn kiểm định T, với α bé cho trước (thường α được lấy bằng 0,05 hoặc 0,01) và với điều kiện H_0 đúng ta có thể tìm được miền W_α sao cho T nhận giá trị trong miền W_α với xác suất bằng α:

$$P\{T \in W_\alpha | H_0\} = \alpha \tag{5.64}$$

Giá trị α được gọi là mức ý nghĩa của tiêu chuẩn kiểm định và miền W_α gọi là *miền bác bỏ giả thiết* H_0 *với mức ý nghĩa* α.

5.3.1.4 Giá trị quan sát của tiêu chuẩn kiểm định

Thực hiện phép thử với mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$ thu được mẫu cụ thể $w=(x_1,x_2,...,x_n)$, thay giá trị này vào thống kê (5.63) ta được giá trị quan sát của tiêu chuẩn kiểm định:

$$T_{qs}=(x_1,x_2,...,x_n,\theta_0) \tag{5.65}$$

5.3.1.5 Quy tắc kiểm định giả thiết thống kê

So sánh giá trị quan sát của tiêu chuẩn kiểm định với miền bác bỏ W_α và kết luận theo quy tắc sau:

1. Nếu $T_{qs}\in W_\alpha$, theo nguyên tắc kiểm định thì H_0 sai, do đó ta bác bỏ H_0 thừa nhận H_1.

2. Nếu $T_{qs}\notin W_\alpha$ thì điều này chưa khẳng định rằng H_0 đúng mà chỉ có nghĩa là qua mẫu cụ thể này chưa khẳng định được là H_0 sai. Do đó ta chỉ có thể nói rằng qua mẫu cụ thể này chưa có cơ sở để bác bỏ H_0 (trên thực tế là thừa nhận H_0).

5.3.1.6 Sai lầm loại một và sai lầm loại hai

Với quy tắc kiểm định như trên có thể mắc hai loại sai lầm sau:

1. Sai lầm loại I: Đó là sai lầm mắc phải khi bác bỏ giả thiết H_0 trong khi H_0 đúng. Ta thấy xác suất mắc sai lầm loại I đúng bằng mức ý nghĩa α. Thật vậy, xác suất ta bác bỏ H_0 bằng xác suất biến cố $\{T\in W_\alpha\}$, do đó khi H_0 đúng thì xác suất này bằng $P\{T\in W_\alpha|H_0\}=\alpha$. Sai lầm loại I sinh ra do kích thước mẫu quá nhỏ, do phương pháp lấy mẫu,v.v…

2. Sai lầm loại II: Đó là sai lầm mắc phải khi thừa nhận giả thiết H_0 trong khi H_0 sai, điều này xảy ra khi giá trị quan sát T_{qs} không thuộc miền bác bỏ W_α trong khi H_1 đúng. Vậy xác suất sai lầm loại II là β xác định như sau:

$$P\{T\notin W_\alpha|H_1\}=\beta \tag{5.66}$$

Xác suất của biến cố đối của sai lầm loại II: $P\{T \in W_\alpha / H_1\}=1-\beta$ gọi là *lực lượng của kiểm định*.

Thực tế / Quyết định	H_0 đúng	H_0 sai
Bác bỏ H_0	Sai lầm loại I Xác suất = α	Quyết định đúng Xác suất = $1-\beta$
Không bác bỏ H_0	Quyết định đúng Xác suất = $1-\alpha$	Sai lầm loại II Xác suất = β

Ta mong muốn tìm một qui tắc kiểm định mà cả hai loại sai lầm trên là cực tiểu. Nhưng không tồn tại kiểm định lý tưởng như vậy, vì nói chung khi giảm sai lầm loại I thì sai lầm loại II tăng và ngược lại. Chẳng hạn nếu lấy $\alpha = 0$ thì sẽ không bác bỏ bất kỳ giả thiết nào, kể cả giả thiết sai, vậy β sẽ đạt cực đại. Mặt khác trong bài toán kiểm định thì giả thiết H_0 là giả thiết quan trọng, do đó sai lầm về nó càng nhỏ càng tốt. Vì vậy các nhà thống kê đưa ra phương pháp sau:

Sau khi ta chọn sai lầm loại I nhỏ ở mức ý nghĩa α, với mẫu kích thước n xác định, ta chọn ra miền bác bỏ W_α sao cho xác suất sai lầm loại II là nhỏ nhất hay lực lượng kiểm định là lớn nhất. Nghĩa là cần tìm miền bác bỏ W_α thỏa mãn điều kiện:

$$P\{T \in W_\alpha | H_0\} = \alpha \text{ và } P\{T \in W_\alpha | H_1\} = 1-\beta \to \max$$

Định lý Neymann - Pearson chỉ ra rằng nhiều bài toán quan trọng trong thực tiễn có thể tìm được miền bác bỏ W_α thỏa mãn điều kiện trên.

Việc chọn mức ý nghĩa α bằng bao nhiêu tùy thuộc vào từng trường hợp cụ thể, tùy thuộc vào ý nghĩa của bài toán.

5.3.1.7 Thủ tục kiểm định giả thiết thống kê

Qua nội dung trình bày ở trên ta có thể xây dựng một thủ tục kiểm định giả thiết thống kê bao gồm các bước sau:

a) Phát biểu giả thiết H_0 và đối thiết H_1.

b) Từ tổng thể nghiên cứu lập mẫu ngẫu nhiên kích thước n.

c) Chọn tiêu chuẩn kiểm định T và xác định quy luật phân bố xác suất của T với điều kiện giả thiết H_0 đúng.

d) Với mức ý nghĩa α, xác định miền bác bỏ W_α tốt nhất tùy thuộc vào đối thiết H_1.

e) Từ mẫu cụ thể tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs}.

f) So sánh giá trị quan sát của tiêu chuẩn kiểm định T_{qs} với miền bác bỏ W_α và kết luận.

5.3.2 Kiểm định giả thiết về kỳ vọng của biến ngẫu nhiên có phân bố chuẩn

Giả sử biến ngẫu nhiên gốc X trong tổng thể có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, cần kiểm định kỳ vọng μ. Nếu có cơ sở để giả thiết rằng kỳ vọng μ bằng giá trị μ_0 ta đưa ra giả thiết thống kê: H_0: $\mu = \mu_0$.

Ta xét các trường hợp sau:

5.3.2.1 Trường hợp đã biết phương sai

Giả sử phương sai σ^2 của biến ngẫu nhiên gốc X trong tổng thể có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ đã biết. Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$.

Xét thống kê

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \tag{5.67}$$

Nếu giả thiết H_0 đúng, theo công thức (5.20) thì thống kê $T = \dfrac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma}$ có phân bố chuẩn tắc $\mathbf{N}(0;1)$. Ta xây dựng các miền bác bỏ dựa vào đối thiết H_1.

a) **Bài toán 1**: H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$. Ta nói đây là *bài toán kiểm định hai phía*.

$$\text{Miền bác bỏ: } W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \; ; \; |T| > U_{\alpha/2} \right\} \tag{5.68}$$

b) **Bài toán 2**: H_0: $\mu = \mu_0$; H_1: $\mu > \mu_0$. Đây là *bài toán kiểm định một phía*.

$$\text{Miền bác bỏ: } W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \; ; \; T > U_\alpha \right\} \tag{5.69}$$

c) **Bài toán 3**: H_0: $\mu = \mu_0$; H_1: $\mu < \mu_0$. Đây là *bài toán kiểm định một phía*.

$$\text{Miền bác bỏ: } W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{\sigma} \; ; \; -T > U_\alpha \right\} \tag{5.70}$$

trong đó $U_{\alpha/2}$, U_α lần lượt là giá trị tới hạn mức $\frac{\alpha}{2}$ và mức α của phân bố chuẩn tắc $\mathbf{N}(0;1)$.

Lập mẫu cụ thể $w = (x_1, x_2, ..., x_n)$, tính giá trị quan sát của tiêu chuẩn kiểm định $T_{qs} = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}$ và so sánh với miền bác bỏ W_α để kết luận.

5.3.2.2 Trường hợp chưa biết phương sai, kích thước mẫu $n \geq 30$

Trường hợp phương sai σ^2 chưa biết: Với kích thước n đủ lớn ($n \geq 30$) và giả thiết H_0 đúng, tương tự mục 5.2.6.2 ta có thống kê:

$$T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \tag{5.71}$$

xấp xỉ phân bố chuẩn tắc $\mathbf{N}(0;1)$. Ta xây dựng các miền bác bỏ dựa vào đối thiết H_1.

a) **Bài toán 1**: H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$. Ta nói đây là *bài toán kiểm định hai phía*.

$$\text{Miền bác bỏ: } W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; |T| > U_{\alpha/2} \right\} \qquad (5.72)$$

b) **Bài toán 2**: H_0: $\mu = \mu_0$; H_1: $\mu > \mu_0$. Đây là *bài toán kiểm định một phía*.

$$\text{Miền bác bỏ: } W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; T > U_\alpha \right\} \qquad (5.73)$$

c) **Bài toán 3**: H_0: $\mu = \mu_0$; H_1: $\mu < \mu_0$. Đây là *bài toán kiểm định một phía*.

$$\text{Miền bác bỏ: } W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; -T > U_\alpha \right\} \qquad (5.74)$$

Ví dụ 5.14: Một hãng buôn muốn biết xem phải chăng có sự không ổn định trung bình về lượng hàng bán được trung bình trên một nhân viên bán hàng so với các năm trước (lượng đó bằng 7,4). Một mẫu ngẫu nhiên gồm 40 nhân viên bán hàng được lựa chọn và tìm thấy lượng hàng trung bình của họ là $\overline{x} = 6{,}1$ với độ lệch chuẩn là $s = 2{,}5$. Với mức ý nghĩa α = 1% có thể nói rằng lượng hàng bán được trung bình trên mỗi đầu người có sự thay đổi không?

Giải: Gọi μ là lượng hàng bán được trung bình trên mỗi nhân viên bán hàng của hãng buôn.

Ta kiểm định: Giả thiết H_0: μ = 7,4; đối thiết H_1: $\mu \neq$ 7,4.

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X} - 7{,}4)\sqrt{n}}{S}$

$$\alpha = 0{,}01 \Rightarrow U_{\frac{\alpha}{2}} = 2{,}575$$

Miền bác bỏ: $W_\alpha = \left\{ T = \dfrac{(\overline{X} - 7{,}4)\sqrt{n}}{S} \;;\; |T| > 2{,}575 \right\}$.

Từ mẫu cụ thể ta có: $T_{qs} = \dfrac{(6{,}1 - 7{,}4)\sqrt{40}}{2{,}5} = -3{,}289$.

Với mẫu cụ thể này giá trị quan sát của tiêu chuẩn kiểm định rơi vào miền bác bỏ, vậy ta có thể kết luận rằng số lượng hàng bán được trung bình của mỗi nhân viên bán hàng là có thay đổi.

Ví dụ 5.15: Một công ty có một hệ thống máy tính có thể xử lí 1200 hóa đơn trong một giờ. Công ty mới nhập một hệ thống máy tính mới. Hệ thống này khi chạy kiểm tra trong 40 giờ cho thấy số hóa đơn được xử lí trung bình trong 1 giờ là 1260 với độ lệch chuẩn 215. Với mức ý nghĩa 5% hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?

Giải: Gọi μ là số hóa đơn trung bình mà hệ thống máy tính mới xử lí được trong 1 giờ.

Ta kiểm định: Giả thiết $H_0: \mu = 1200$; đối thiết $H_1: \mu > 1200$.

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X} - 1200)\sqrt{n}}{S}$

$$\alpha = 0{,}05 \Rightarrow U_\alpha = 1{,}64$$

Miền bác bỏ: $W_\alpha = \left\{ T = \dfrac{(\overline{X} - 1200)\sqrt{n}}{S} \;;\; T > 1{,}64 \right\}$.

Thay giá trị cụ thể của mẫu vào công thức (5.61) ta được $T_{qs} = \dfrac{(1260 - 1200)\sqrt{40}}{215} = 1{,}76$. Với mẫu cụ thể này giá trị quan sát của tiêu chuẩn kiểm định rơi vào miền bác bỏ, vậy ta có thể kết luận hệ thống máy tính mới tốt hơn hệ thống cũ.

5.3.2.3 Trường hợp chưa biết phương sai, kích thước mẫu n<30

Giả sử giả thiết H_0 đúng, xét thống kê:

$$T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \tag{5.75}$$

theo công thức (5.23) thống kê T có phân bố Student $n-1$ bậc tự do. Ta xây dựng các miền bác bỏ dựa vào đối thiết H_1.

a) **Bài toán 1**: H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$.

Miền bác bỏ: $W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; |T| > t_{\alpha/2}(n-1) \right\}$ (5.76)

trong đó $t_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\frac{\alpha}{2}$ của phân bố Student $n-1$ bậc tự do.

b) **Bài toán 2**: H_0: $\mu = \mu_0$; H_1: $\mu > \mu_0$.

Miền bác bỏ: $W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; T > t_{\alpha}(n-1) \right\}$ (5.77)

trong đó $t_{\alpha}(n-1)$ là giá trị tới hạn mức α của phân bố Student $n-1$ bậc tự do.

c) **Bài toán 3**: H_0: $\mu = \mu_0$; H_1: $\mu < \mu_0$.

Miền bác bỏ: $W_\alpha = \left\{ T = \frac{(\overline{X} - \mu_0)\sqrt{n}}{S} \; ; \; -T > t_{\alpha}(n-1) \right\}$ (5.78)

Ví dụ 5.16: Một công ty sản xuất hạt giống tuyên bố rằng một loại giống mới của họ có năng suất trung bình là 21,5 tạ/ha. Gieo thử hạt giống mới này tại 16 vườn thí nghiệm và thu được kết quả:

19,2; 18,7; 22,4; 20,3; 16,8; 25,1; 17,0; 15,8;
21,0; 18,6; 23,7; 24,1; 23,4; 19,8; 21,7; 18,9.

Dựa vào kết quả này hãy xác nhận xem quảng cáo của công ty có đúng không. Mức ý nghĩa được lựa chọn là α = 0,05. Biết rằng năng suất giống cây trồng là một biến ngẫu nhiên có phân bố chuẩn $\mathbf{N}(\mu; \sigma^2)$.

Giải: Gọi μ là năng suất trung bình của loại giống mới.

Ta cần kiểm định giả thiết H_0: $\mu = 21{,}5$; đối thiết H_1: $\mu \neq 21{,}5$.

Tiêu chuẩn kiểm định: $T = \dfrac{(\overline{X} - 21{,}5)\sqrt{n}}{S}$.

Tra bảng ta tính được giá trị tới hạn mức $\dfrac{\alpha}{2} = 0{,}025$ của phân bố Student 15 bậc tự do là 2,131. Do đó miền bác bỏ:

$$W_\alpha = \left\{ T = \frac{(\overline{X} - 21{,}5)\sqrt{n}}{S} \; ; \; |T| > 2{,}131 \right\}.$$

Từ mẫu cụ thể trên tính được:

$$\overline{x} = 20{,}406\,, \quad s = 3{,}038$$

$$\Rightarrow T_{qs} = \frac{(20{,}406 - 21{,}5)\sqrt{16}}{3{,}038} = -1{,}44\,.$$

Vì $|T_{qs}| = 1{,}44 < 2{,}131$ nên chưa có cơ sở để bác bỏ H_0. Có nghĩa là với số liệu này thì có thể chấp nhận lời quảng cáo của công ty.

5.3.3 Kiểm định giả thiết về tần suất

Giả sử ta để ý đến một đặc trưng A nào đó mà mỗi cá thể của tổng thể có thể có tính chất này hoặc không. Gọi p là tần suất có đặc trưng A của tổng thể, như đã thấy trong mục 5.1.5.5 và mục 5.2.4, dấu hiệu nghiên cứu này là một biến ngẫu nhiên X có phân bố Bernoulli tham số p với kỳ vọng bằng p. Nếu p chưa biết, song có cơ sở để giả thiết rằng giá trị này bằng p_0. Ta kiểm định giả thiết H_0: $p = p_0$.

Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n, gọi f là tần suất mẫu (công thức (5.13), (5.15)). Với giả thiết H_0 đúng, xét thống kê

$$T = \frac{(f - p_0)\sqrt{n}}{\sqrt{p_0(1 - p_0)}} \tag{5.79}$$

Áp dụng định lý giới hạn trung tâm (công thức (5.25)) thì khi n đủ lớn thống kê (5.79) xấp xỉ phân bố chuẩn tắc $\mathbf{N}(0;1)$. Trong thực tế khi:

$$\begin{cases} np_0 > 5 \\ n(1-p_0) > 5 \end{cases} \tag{5.80}$$

thì có thể xem thống kê (5.79) có phân bố chuẩn tắc $\mathbf{N}(0;1)$. Do đó với mức ý nghĩa α và tùy thuộc đối thiết H_1 ta có thể xây dựng các miền bác bỏ tương ứng.

a) H_0: $p = p_0$; H_1: $p \neq p_0$.

$$W_\alpha = \left\{ T = \frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \; ; \; |T| > U_{\frac{\alpha}{2}} \right\} \tag{5.81}$$

b) H_0: $p = p_0$; H_1: $p > p_0$.

$$W_\alpha = \left\{ T = \frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \; ; \; T > U_\alpha \right\} \tag{5.82}$$

c) H_0: $p = p_0$; H_1: $p < p_0$.

$$W_\alpha = \left\{ T = \frac{(f-p_0)\sqrt{n}}{\sqrt{p_0(1-p_0)}} \; ; \; -T > U_\alpha \right\} \tag{5.83}$$

Với mẫu cụ thể tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs}, so sánh với W_α và kết luận.

Ví dụ 5.17: Một đảng chính trị trong một cuộc bầu cử tổng thống ở nước nọ tuyên bố rằng có 45% cử tri sẽ bỏ phiếu cho ứng cử viên A của đảng đó.

Chọn ngẫu nhiên 2000 cử tri để thăm dò ý kiến và cho thấy có 862 cử tri tuyên bố sẽ bỏ phiếu cho A.

Với mức α = 5%, hãy kiểm định xem dự đoán của đảng trên có đúng không.

Giải: Gọi p là tỉ lệ cử tri sẽ bỏ phiếu cho ứng cử viên A.

Ta cần kiểm định: Giả thiết H_0: $p = 0{,}45$; Đối thiết H_1: $p \neq 0{,}45$.

(Bởi vì ta không có cơ sở nào để cho rằng dự đoán của đảng trên là cao hơn 0,45 hay thấp hơn 0,45).

Vì rằng điều kiện $\begin{cases} np_0 = 2000 \cdot 0{,}45 = 900 > 5 \\ n(1-p_0) = 2000 \cdot 0{,}55 = 1100 > 5 \end{cases}$ thỏa mãn nên có thể chọn tiêu chuẩn kiểm định theo công thức (5.79).

Thay mẫu cụ thể với $f = \dfrac{862}{2000} = 0{,}431$ ta được giá trị quan sát của tiêu chuẩn kiểm định $T_{qs} = \dfrac{(0{,}431 - 0{,}45)\sqrt{2000}}{\sqrt{0{,}45 \cdot 0{,}55}} = -1{,}708$.

Với mức ý nghĩa $\alpha = 0{,}05 \Rightarrow U_{\frac{\alpha}{2}} = 1{,}96$. Ta thấy $|T_{qs}| < 1{,}96$. Vậy không có cơ sở để bác bỏ H_0.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

5.1 Mẫu ngẫu nhiên kích thước n về dấu hiệu nghiên cứu X là một dãy gồm n biến ngẫu nhiên: $X_1, X_2, ..., X_n$ độc lập cùng phân bố với X.

Đúng ☐ Sai ☐ .

5.2 Một thống kê của mẫu ngẫu nhiên là con số cụ thể về dấu hiệu nghiên cứu.

Đúng ☐ Sai ☐ .

5.3 Trung bình mẫu của dấu hiệu nghiên cứu có phân bố chuẩn cũng có phân bố chuẩn.

Đúng ☐ Sai ☐ .

5.4 Một thống kê của mẫu là một hàm của các biến ngẫu nhiên thành phần của mẫu do đó cũng là một biến ngẫu nhiên .

Đúng ☐ Sai ☐ .

5.5 Trung bình mẫu là ước lượng vững và hiệu quả của kỳ vọng của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐ .

5.6 Có thể tìm được ước lượng không chệch của θ có phương sai nhỏ hơn đại lượng $\dfrac{1}{n\mathrm{E}\left(\dfrac{\partial\left(\ln f(X,\theta)\right)}{\partial\theta}\right)^2}$.

Đúng ☐ Sai ☐ .

5.7 Tổng của hai ước lượng không chệch là một ước lượng không chệch.

Đúng ☐ Sai ☐ .

5.8 Phương sai mẫu hiệu chỉnh S^2 là ước lượng vững không chệch của phương sai của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐ .

5.9 Hai đầu mút của khoảng tin cậy là hai thống kê của mẫu.

Đúng ☐ Sai ☐ .

5.10 Muốn tìm khoảng tin cậy cho tham số μ của biến ngẫu nhiên gốc có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì kích thước mẫu n phải lớn hơn 30.

Đúng ☐ Sai ☐ .

5.11 Giả thiết thống kê là giả thiết do nhà thống kê đặt ra cho mẫu ngẫu nhiên.

Đúng ☐ Sai ☐ .

5.12 Bác bỏ giả thiết dẫn đến chấp nhận đối thiết và ngược lại do đó đối thiết là phủ định của giả thiết.

Đúng ☐ Sai ☐ .

5.13 Qui tắc kiểm định dựa trên nguyên lý xác suất nhỏ và phép chứng minh phản chứng.

Đúng ☐ Sai ☐ .

5.14 Sai lầm loại 1 là sai lầm gặp phải khi thực tế giả thiết đúng nhưng ta bác bỏ.

Đúng ☐ Sai ☐ .

5.15 Sai lầm loại 2 luôn luôn lớn hơn sai lầm loại 1.

Đúng ☐ Sai ☐ .

5.16 Miền bác bỏ là miền có xác suất rất bé nên ta có thể bỏ qua trong mọi phép kiểm định.

Đúng ☐ Sai ☐ .

5.17 Khi xây dựng tiêu chuẩn kiểm định T ta luôn giả sử rằng giả thiết H_0 sai vì giả thiết H_0 là điều ta nghi ngờ muốn bác bỏ.

Đúng ☐ Sai ☐ .

5.18 Kiểm định hai phía là kiểm định đối với những tham số có thể nhận giá trị âm dương bất kỳ, còn kiểm định một phía khi tham số cần kiểm định chỉ nhận giá trị dương hoặc âm.

Đúng ☐ Sai ☐ .

5.19 Từ tổng thể có dấu hiệu nghiên cứu X có bảng phân bố xác suất sau:

X	0	1
P	0,5	0,5

lập mẫu ngẫu nhiên kích thước $n=10$. Tính xác suất để trung bình mẫu của mẫu ngẫu nhiên này nhận giá trị 0,5.

5.20 Giả sử biến ngẫu nhiên gốc có phân bố chuẩn $\mathbf{N}(20;1)$. Chọn mẫu ngẫu nhiên kích thước $n=100$. Hãy tính xác suất để trung bình mẫu $\overline{X}$ nằm trong khoảng: $19{,}8 < \overline{X} < 20{,}2$.

5.21 Một mẫu cụ thể của biến ngẫu nhiên X như sau:

2 ; 3 ; 2 ; 4 ; 1 ; 4 ; 2 ; 2 ; 3 ; 1 ($n=10$).

a) Lập bảng phân bố tần suất.

b) Xây dựng hàm phân bố thực nghiệm.

Tính $\overline{x}$, s^2, s.

5.22 Giả sử $\Theta_n = T(X_1, X_2, ..., X_n)$ là một ước lượng của θ thỏa mãn: $\lim\limits_{n\to\infty} \mathrm{E}(\Theta_n) = \theta$ và $\lim\limits_{n\to\infty} \mathrm{D}(\Theta_n) = 0$.

Chứng minh Θ_n là một ước lượng vững của θ.

5.23 Giả sử $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên của biến ngẫu nhiên gốc X có phân bố Poisson tham số λ. Hãy ước lượng hợp lý cực đại tham số λ.

5.24 Giả sử $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên của biến ngẫu nhiên gốc X có phân bố mũ tham số λ. Hãy ước lượng hợp lý cực đại tham số λ.

5.25 Trong đợt vận động bầu cử tổng thống ở một nước nọ, người ta phỏng vấn ngẫu nhiên 2000 cử tri thì được biết có 1082 người trong số đó sẽ bỏ phiếu cho ứng cử viên A. Với độ tin cậy 98% tối thiểu ứng cử viên A sẽ chiếm được bao nhiêu % số phiếu bầu?

5.26 Để xác định sản lượng khai thác điện thoại của đơn vị mình, một đơn vị đã tiến hành thống kê ngẫu nhiên 35 ngày và thu được kết quả sau với đơn vị 100.000 phút/ngày:

0,84	0,96	1,02	1,08	0,88	0,80	0,91	0,97	1,07	0,98
1,04	1,13	0,87	0,82	1,01	0,93	1,03	1,10	0,97	1,05
0,83	0,76	0,95	1,15	1,00	1,05	1,14	0,89	0,81	0,95
1,20	1,16	1,24	0,79	0,77					

Tìm khoảng tin cậy 95% cho sản lượng điện thoại trung bình mỗi ngày.

5.27 Muốn ước lượng số cá trong hồ, người ta bắt 2000 con cá trong hồ đánh dấu rồi thả lại xuống hồ. Sau đó bắt lại 400 con và thấy có 53 con có dấu. Hãy ước lượng số cá trong hồ với độ tin cậy là 0,95.

5.28 Để xác định chiều cao trung bình của các cây con trong một vườn ươm người ta tiến hành đo ngẫu nhiên 40 cây. Kết quả đo được như sau:

Khoảng chiều cao (cm)	16,5-17	17-17,5	17,5-18	18-18,5	18,5-19	19-19,5
Số cây tương ứng	3	5	11	12	6	3

a) Tìm khoảng tin cậy 90% cho chiều cao trung bình của vườn cây con.

b) Nếu muốn khoảng ước lượng có độ chính xác $\varepsilon = 0,1$ thì cần lấy mẫu bao nhiêu cây.

5.29 Trọng lượng của một loại sản phẩm A là một biến ngẫu nhiên có phân bố theo quy luật chuẩn với độ lệch chuẩn là 1 gam. Cân thử 27 bao loại này ta thu được kết quả:

Trọng lượng (gram)	47,5-48,5	48,5-49,5	49,5-50,5	50,5-51,5	51,5-52,5
Số bao tương ứng	3	6	15	2	1

a) Tìm khoảng tin cậy 95% của trọng lượng trung bình của loại sản phẩm trên.

b) Nếu muốn độ chính xác $\varepsilon = 0,1$ thì kích thước mẫu cần thiết là bao nhiêu.

5.30 Người ta đo một đại lượng không đổi 25 lần bằng một dụng cụ đo không có sai số hệ thống và sai số đo trung bình bằng 0. Giả sử sai số của phép đo tuân theo quy luật phân bố chuẩn và phương sai mẫu đo được bằng 0,5. Hãy xác định khoảng tin cậy 95% cho phương sai của sai số đo.

5.31 Trọng lượng đóng bao của một loại sản phẩm X là biến ngẫu nhiên có phân bố theo quy luật chuẩn với trọng lượng trung bình theo quy định là 100kg. Nghi ngờ sản phẩm bị đóng thiếu, người ta cân thử 29 bao loại này thu được kết quả:

Trọng lượng (kg)	98,0 -98,5	98,5– 99,0	99,0 - 99,5	99,5 - 100	100 -100,5	100,5-101
Số bao tương ứng	2	6	10	7	3	1

Với mức ý nghĩa $\alpha = 0,025$ hãy kết luận về điều nghi ngờ nói trên.

5.32 Định mức thời gian hoàn thành sản phẩm là 14 phút. Liệu có cần thay đổi định mức không, nếu theo dõi thời gian hoàn thành sản phẩm ở 250 công nhân ta thu được kết quả như sau:

X (phút)	10 - 12	12 - 14	14 - 16	16 - 18	18 - 20
Số công nhân	20	60	100	40	30

Với mức ý nghĩa $\alpha = 0,025$ hãy kết luận về ý định nói trên.

5.33 Mức hao phí xăng của một loại ô tô chạy từ A đến B là một biến ngẫu nhiên có quy luật chuẩn với kỳ vọng 50 lít. Đoạn đường được sửa chữa lại. Người ta cho rằng mức hao phí xăng trung bình giảm xuống. Quan sát 28 ô tô cùng loại thu được:

X hao phí (lít)	48,5 - 49,0	49,0 - 49,5	49,5 - 50,0	50,0 - 50,5	50,5-51
Số ô tô tương ứng	4	10	9	3	2

Với mức ý nghĩa $\alpha = 0,025$ hãy kết luận về điều nghi ngờ nói trên.

5.34 Một công ty có một hệ thống máy tính có thể xử lý 1300 hoá đơn trong 1 giờ. Công ty mới nhập một hệ thống máy tính mới, hệ thống này chạy kiểm tra trong 40 giờ cho thấy số hoá đơn xử lý trung bình trong 1 giờ là 1378 với độ lệch tiêu chuẩn 215. Với mức ý nghĩa 2,5% hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?

Chương 6

QUÁ TRÌNH NGẪU NHIÊN, CHUỖI MARKOV

GIỚI THIỆU

Hầu hết các hiện tượng xảy ra trong tự nhiên và xã hội đều có tính chất ngẫu nhiên, đó là sự phản ánh của các mối ràng buộc phức tạp mà ta không biết trước được. Trong các chương trước chúng ta đã tìm hiểu khái niệm biến ngẫu nhiên, véc tơ ngẫu nhiên, đó là các biến nhận giá trị nào đó phụ thuộc vào các yếu tố ngẫu nhiên. Khi họ các biến ngẫu nhiên phụ thuộc vào thời gian ta có quá trình ngẫu nhiên.

Lý thuyết quá trình ngẫu nhiên lần đầu tiên được nghiên cứu liên quan đến bài toán dao động và nhiễu của các hệ vật lý. Quá trình ngẫu nhiên là một mô hình toán học của quá trình thực nghiệm mà sự phát triển bị chi phối bởi các quy luật xác suất. Quá trình ngẫu nhiên cung cấp những mô hình hữu ích để nghiên cứu nhiều lĩnh vực khác nhau như vật lý thống kê, viễn thông, điều khiển, phân tích chuỗi thời gian, sự tăng trưởng dân số và các ngành khoa học quản lý.

Các tín hiệu video, tín hiệu thoại, dữ liệu máy tính, nhiễu của một hệ thống viễn thông, nhiễu điện trong các thiết bị điện, số khách hàng đến một điểm phục vụ, chỉ số chứng khoán trong thị trường chứng khoán… là các quá trình ngẫu nhiên.

Quá trình ngẫu nhiên có nhiều ứng dụng trong viễn thông là quá trình Markov (quá trình không nhớ - memoryless) và quá trình dừng.

Chuỗi Markov là một quá trình Markov có không gian trạng thái rời rạc, thời gian rời rạc và thuần nhất. Chuỗi Markov thường gặp trong bài toán chuyển mạch của hệ thống viễn thông.

Quá trình Poisson là một ví dụ về chuỗi Markov với thời gian liên tục. Quá trình Poisson $X(t)$ mô tả quá trình đếm số lần xuất hiện một biến cố A nào đó cho đến thời điểm t. Quá trình Poisson được ứng dụng nhiều trong viễn thông, liên quan đến bài toán truyền tín hiệu, các hệ phục vụ, bài toán chuyển mạch ...

Tín hiệu viễn thông, nhiễu không có tính Markov. Các quá trình này, quá khứ của nó có ảnh hưởng lớn đến sự tiến triển của quá trình trong tương lai. Tuy nhiên hàm trung bình không thay đổi và hàm tương quan thuần nhất theo thời gian, đó là quá trình dừng. Khi các quá trình dừng biểu diễn các tín hiệu hoặc nhiễu thì biến đổi Fourier của hàm tương quan của quá trình là hàm mật độ phổ công suất của tín hiệu hoặc nhiễu này.

Một trong những bài toán quan trọng của lý thuyết chuyển mạch là vấn đề xung đột thông tin, nghẽn mạch hoặc rớt cuộc gọi. Lý thuyết quá trình sắp hàng (Queueing theory) xác định và tìm các phương án tối ưu để hệ thống phục vụ tốt nhất.

Trong chương này ta chỉ nghiên cứu một cách khái quát khái niệm quá trình ngẫu nhiên và chuỗi Markov với thời gian rời rạc thuần nhất.

Để học tốt chương này học viên cần nắm vững khái niệm xác suất, xác suất có điều kiện, công thức xác suất đầy đủ, biến ngẫu nhiên và các kiến thức đại số tuyến tính như ma trận, hệ phương trình tuyến tính.

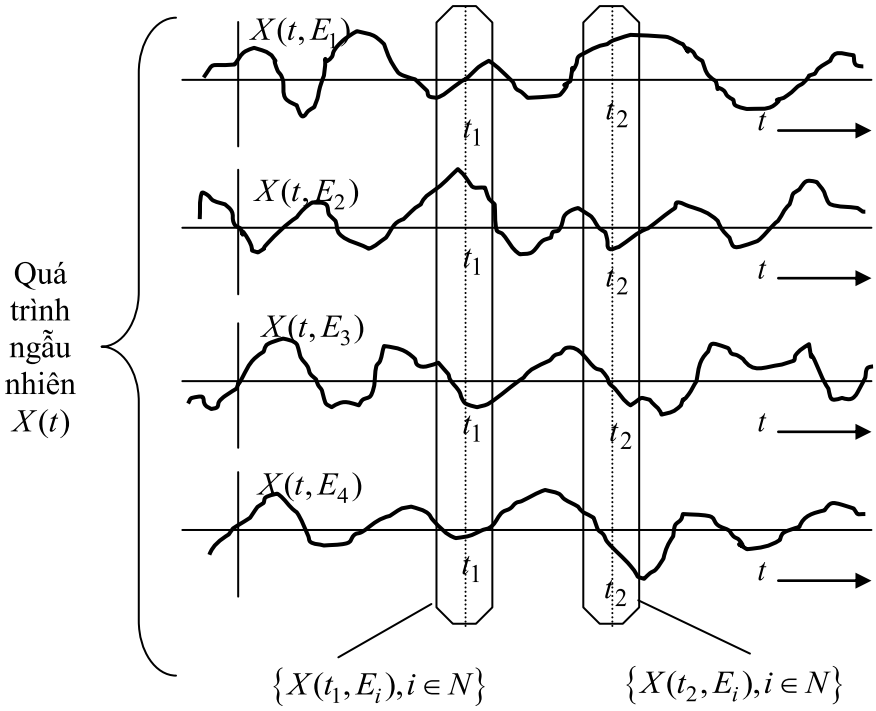
6.1 KHÁI NIỆM VÀ PHÂN LOẠI QUÁ TRÌNH NGẪU NHIÊN

6.1.1 Khái niệm quá trình ngẫu nhiên

Các tín hiệu của các hệ thống thông tin là các tín hiệu ngẫu nhiên vì ngoài thành phần mang tin còn có sự tác động của giao thoa ngẫu nhiên và nhiễu của thiết bị.

Giả sử một tín hiệu nào đó mà tại mỗi thời điểm t nhận các giá trị phụ thuộc hệ các biến cố $\{E_i, i \in N\}$ của phép thử. Tín hiệu này nhận giá trị là $X(t, E_i)$ tại thời điểm t và khi biến cố E_i xảy ra. Như vậy $\{X(t, E_i)\}$ là một hàm mẫu của quá trình ngẫu nhiên $X(t)$. Quá

trình ngẫu nhiên $X(t)$ vừa phụ thuộc thời gian t, vừa phụ thuộc yếu tố ngẫu nhiên E_i.



Hình 6.1: Mô hình quá trình ngẫu nhiên

Tổng quát một quá trình ngẫu nhiên là một họ các biến ngẫu nhiên $\{X(t,\omega); t \in T\}$ *xác định trong cùng một phép thử. Các quá trình này vừa phụ thuộc vào thời gian* t *và khi cố định tham số* t *thì* $X(t,\omega)$ *là biến ngẫu nhiên theo* ω. *Khi cố định* ω *ta được hàm theo thời gian* t *và được gọi là* ***hàm mẫu*** *hoặc một thể hiện của quá trình ngẫu nhiên. Tập chỉ số* T *thường biểu diễn tham số thời gian.*

Do tác động của các yếu tố ngẫu nhiên nên một tín hiệu $\{X(t,\omega); t \in T\}$ được truyền đi là một quá trình ngẫu nhiên. Tín hiệu cụ thể nhận được là hàm mẫu (một thể hiện) của quá trình ngẫu nhiên $\{X(t,\omega); t \in T\}$.

Để đơn giản trong cách viết người ta ký hiệu quá trình ngẫu nhiên $\{X(t); t \in T\}$ thay cho $\{X(t, \omega); t \in T\}$, hàm mẫu tương ứng được ký hiệu $\{x(t); t \in T\}$.

6.1.2 Phân loại quá trình ngẫu nhiên

Có thể phân loại các quá trình ngẫu nhiên theo các đặc trưng sau:

- Không gian trạng thái;
- Tập chỉ số thời gian T;
- Quan hệ độc lập, quy luật phân bố xác suất của các biến ngẫu nhiên $X(t)$.

6.1.2.1 Phân loại quá trình ngẫu nhiên theo tập trạng thái E

Ta ký hiệu E là tập các giá trị của $X(t)$ và gọi là không gian trạng thái của quá trình, mỗi giá trị của $X(t)$ được gọi là một trạng thái.

- Nếu E là tập đếm được thì $\{X(t); t \in T\}$ gọi là quá trình có trạng thái rời rạc.
- Nếu E là 1 khoảng của tập số thực 3 thì $\{X(t); t \in T\}$ được gọi là quá trình thực hoặc quá trình trạng thái liên tục.
- Nếu E là tập con của tập số phức $\forall$ thì $\{X(t); t \in T\}$ là quá trình trạng thái phức.
- Nếu $E \subset 3^k$ thì $\{X(t); t \in T\}$ là quá trình trạng thái k-véc tơ.

6.1.2.2 Phân loại quá trình ngẫu nhiên theo tập các chỉ số T

- Nếu $T \subset 9$ thì quá trình $\{X(t); t \in T\}$ được gọi là quá trình có thời gian rời rạc hoặc tham số rời rạc. Trường hợp này ta ký hiệu X_n thay cho $X(t)$ và gọi là *một dãy ngẫu nhiên*.
- Nếu $T = [0; \infty)$ hoặc $T = 3$ thì $\{X(t); t \in T\}$ được gọi là quá trình có thời gian liên tục.

6.1.2.3 Phân loại theo các tính chất xác suất của quá trình ngẫu nhiên

a) Quá trình độc lập

Quá trình $\{X(t); t \in T\}$ được gọi là quá trình độc lập nếu với mọi thời điểm $t_1 < t_2 < ... < t_n$ thì các biến ngẫu nhiên sau là độc lập

$$X(t_1), X(t_2), ..., X(t_n) \tag{6.1}$$

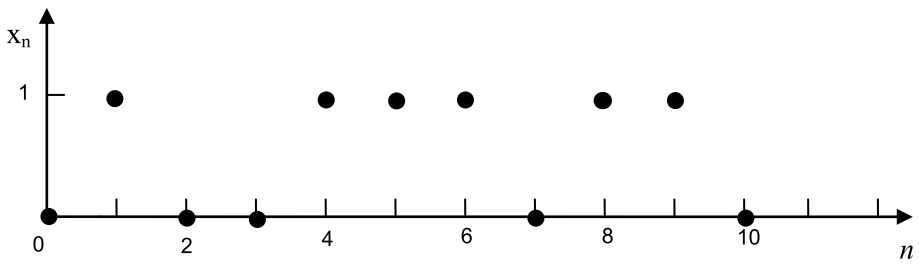
Ví dụ 6.1: Giả sử $X_1, X_2, ...$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli với xác suất $P\{X_n = 1\} = p$, $P\{X_n = 0\} = q = 1 - p$ với mọi n. Khi đó $\{X_n, n \geq 1\}$ là một quá trình ngẫu nhiên gọi là quá trình Bernoulli.

Quá trình Bernoulli là quá trình độc lập có không gian trạng thái rời rạc $E = \{0,1\}$, thời gian rời rạc $T = \{1,2,...\}$.

Một ví dụ mô phỏng quá trình Bernoulli có thể nhận được bằng cách gieo đồng xu liên tiếp. Nếu mặt sấp xuất hiện ta gán giá trị 1, nếu mặt ngửa xuất hiện ta gán giá trị 0, các giá trị 0, 1 là trạng thái của quá trình. Chẳng hạn ta có một dãy mẫu của quá trình Bernoulli như sau:

n	1	2	3	4	5	6	7	8	9	10	...
Mặt xuất hiện	S	N	N	S	S	S	N	S	S	N	...
x_n	1	0	0	1	1	1	0	1	1	0	...

Dãy mẫu $\{x_n, n \geq 1\}$ nhận được ở trên được minh họa trong hình sau:



Hình 6.2: Hàm mẫu của quá trình Bernoulli

b) Quá trình có gia số độc lập

Quá trình $\{X(t); t \in T\}$ được gọi là quá trình gia số độc lập nếu các gia số của quá trình trong các khoảng thời gian rời nhau là các biến ngẫu nhiên độc lập. Tức là với mọi cách chọn các thời điểm $t_1 < t_2 < ... < t_n$ thì các biến ngẫu nhiên sau là độc lập:

$$X(t_2) - X(t_1), X(t_3) - X(t_2), ..., X(t_n) - X(t_{n-1}). \qquad (6.2)$$

Đặc biệt với quá trình thời gian rời rạc $\{X_n\}$ thì tính chất gia số độc lập dẫn đến dãy các biến ngẫu nhiên $Z_0 = X_0$, $Z_i = X_i - X_{i-1}$; $i = 1, 2, ...$ là độc lập. Ngoài ra nếu ta biết luật phân bố của từng biến ngẫu nhiên $Z_0, Z_1, ...$ thì ta biết được luật phân bố của mọi X_i, $i = 0, 1, ...$

Thật vậy, điều này được suy từ công thức (3.50), (3.51a), (3.51b) và

$$X_i = Z_0 + Z_1 + ... + Z_i$$

c) Quá trình gia số độc lập dừng

Quá trình gia số độc lập $\{X(t); t \in T\}$ được gọi là quá trình gia số độc lập dừng nếu:

$\forall s, t, s < t$; $\forall h \geq 0$: $X(t) - X(s)$ và $X(t+h) - X(s+h)$ có cùng phân bố (6.3)

Quá trình Poisson, quá trình Wiener là hai ví dụ của quá trình gia số độc lập dừng.

d) Quá trình Martingal

Quá trình $\{X(t); t \in T\}$ được gọi là quá trình Martingal nếu với bất kỳ $t_1 < t_2 < ... < t_n < t_{n+1}$ và $a_1, a_2, ..., a_n$ thì:

$$\mathrm{E}\left[X(t_{n+1}) \middle| X(t_1) = a_1, ..., X(t_n) = a_n \right] = a_n \qquad (6.4)$$

Martingal có thể xem như là mô hình mô tả trò chơi may rủi, trong đó $X(t)$ là số tiền của người chơi ở thời điểm t. Tính chất

Martingal nói rằng *số tiền trung bình* của người chơi sẽ có ở thời điểm t_{n+1} bằng số tiền anh ta có ở thời điểm t_n và *không phụ thuộc vào những gì anh ta có trước đó trong quá khứ.*

Nếu $\{X(t); t \geq 0\}$ là quá trình gia số độc lập với kỳ vọng bằng 0 thì $\{X(t); t \geq 0\}$ là quá trình Martingal với thời gian liên tục.

e) Quá trình Markov

Quá trình $\{X(t); t \in T\}$ được gọi là quá trình Markov nếu:

Với mọi thời điểm $t_1 < t_2 < ... < t_n$, với mọi giá trị $a_1, a_2, ..., a_n$ cho trước, với mọi thời điểm $t > t_n$ và với mọi a ta có:

$$P\{X(t) \leq a \,|\, X(t_1) = a_1, ..., X(t_n) = a_n\} = P\{X(t) \leq a \,|\, X(t_n) = a_n\} \tag{6.5}$$

Nghĩa là phân bố xác suất trong tương lai chỉ phụ thuộc hiện tại và độc lập với quá khứ. Nói cách khác *quá trình Markov mô tả các hệ không nhớ* (memoryless).

Với mọi $t > s$; với mọi tập giá trị $A \subset \mathbb{R}$ và giá trị b ta ký hiệu:

$$p(s, b; t, A) = P\{X(t) \in A \,|\, X(s) = b\} \tag{6.6}$$

và gọi là hàm xác suất chuyển từ thời điểm s đến thời điểm t.

Như vậy công thức (6.5) được viết lại:

$$P\{X(t) \leq a \,|\, X(t_1) = a_1, ..., X(t_n) = a_n\} = p(t_n, a_n; t, A),$$

trong đó $A = (-\infty, a]$. (6.7)

Quá trình Markov với không gian trạng thái rời rạc được gọi là chuỗi Markov (hay xích Markov, Markov chains). Chuỗi Markov với thời gian rời rạc và thuần nhất được xét trong mục tiếp theo.

Quy luật phân bố xác suất của biến ngẫu nhiên rời rạc được xét qua hàm khối lượng xác suất, vì vậy tính chất Markov – công thức

(6.5) đối với chuỗi Markov $\{X_n; n = 0,1,2,...\}$ với thời gian rời rạc được viết lại như sau:

$$P\{X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, ..., X_n = i_n\} = P\{X_{n+1} = j \mid X_n = i_n\}$$

với $i_0, i_1, ..., i_n, j \in E$. (6.8)

f) Quá trình dừng (stationary)

Quá trình $\{X(t); t \in T\}$, $T =$ 3 , 3 $_+$, 9 hoặc Đ được gọi là:

- *Dừng theo nghĩa chặt* (strictly stationary) nếu với mọi $h > 0$, với mọi $t_1, t_2, ..., t_n \in T$ thì hai véc tơ ngẫu nhiên

$$(X(t_1 + h), X(t_2 + h), ..., X(t_n + h))$$

và $(X(t_1), X(t_2), ..., X(t_n))$ có cùng phân bố xác suất.

Nói riêng mọi $X(t)$ có cùng phân bố.

Quá trình dừng theo nghĩa chặt ít gặp trong thực tế và người ta mở rộng khái niệm quá trình dừng theo nghĩa sau.

- *Dừng theo nghĩa rộng hay dừng hiệp phương sai* (wide sense stationary or covariance stationary) nếu

(i) $\mathrm{E}\, X(t) = m =$ const

(ii) Với mọi t, $\mathrm{E}[X(t), X(t+\tau)]$ chỉ phụ thuộc τ.

Đặt: $K_x(\tau) = \mathrm{E}[X(t), X(t+\tau)]$ (6.9)

và gọi là *hàm tự tương quan* của quá trình $\{X(t); t \in T\}$.

6.2 CHUỖI MARKOV

Chuỗi Markov là quá trình Markov $\{X(t); t \in T\}$ có không gian trạng thái E đếm được.

Tuỳ theo tập chỉ số $T = \{0,1,2,...\}$ hoặc $T = (0;\infty)$ ta có tương ứng chuỗi Markov với thời gian rời rạc hoặc liên tục.

Với chuỗi Markov công thức xác suất chuyển (6.6) được viết cụ thể:

$$p(s,i;t,j) = P\{X(t) = j \,|\, X(s) = i\},\ t > s; i, j \in E \qquad (6.10)$$

Nếu xác suất chuyển (6.10) chỉ phụ thuộc vào $t - s$ nghĩa là:

$$p(s,i;t,j) = p(s+h,i;t+h,j) \qquad (6.11)$$

với mọi h, thì ta nói quá trình là thuần nhất theo thời gian.

6.2.1 Chuỗi Markov với thời gian rời rạc thuần nhất

Định nghĩa 6.1: *Quá trình* $\{X_n,\ n = 0,1,2,...\}$ *với thời gian rời rạc được gọi là chuỗi Markov thời gian rời rạc thuần nhất nếu thỏa mãn hai điều kiện sau*

i) Không gian trạng thái E của mọi X_n *là tập đếm được.*

ii) Hàm xác suất chuyển là thuần nhất theo thời gian, tức là thoả mãn (6.11).

Từ giờ ta chỉ xét chuỗi Markov với thời gian rời rạc thuần nhất và ta gọi tắt là chuỗi Markov.

6.2.2 Ma trận xác suất chuyển

Giả sử $\{X_n,\ n = 0,1,2,...\}$ là chuỗi Markov thời gian rời rạc có không gian trạng thái E. Các phần tử của E được ký hiệu $i, j, k...$

Với mọi $i, j \in E$; đặt

$$p_{ij} = P\{X_{n+1} = j \,|\, X_n = i\} = P\{X_1 = j \,|\, X_0 = i\} \qquad (6.12)$$

không phụ thuộc vào n. Đó là xác suất để từ trạng thái i sau một bước sẽ chuyển thành trạng thái j.

Định nghĩa 6.2: *Ma trận* $P = \left[p_{ij} \right]$ *với* p_{ij} *xác định theo (6.12) được gọi là ma trận xác suất chuyển hay ma trận xác suất chuyển sau 1 bước của chuỗi Markov* $\{X_n,\ n = 0,1,2,...\}$.

Các phần tử p_{ij} của ma trận xác suất chuyển thỏa mãn điều kiện

$$p_{ij} \geq 0;\ \sum_{j \in E} p_{ij} = 1,\ \forall i \in E \tag{6.13}$$

- Nếu tập trạng thái E vô hạn thì ma trận xác suất chuyển có vô số hàng, vô số cột và tổng trong công thức (6.13) là tổng của một chuỗi số dương.

- Nếu tập trạng thái E hữu hạn, chẳng hạn $E = \{1, 2, ..., m\}$ thì ma trận xác suất chuyển và công thức (6.13) được viết dưới dạng

$$P = \left[p_{ij} \right] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix} \tag{6.14}$$

$$p_{ij} \geq 0;\ \sum_{j=1}^{m} p_{ij} = 1,\ \forall i = 1, ..., m \tag{6.15}$$

Ma trận vuông thỏa mãn điều kiện (6.15) được gọi là *Ma trận Markov* hoặc ma trận ngẫu nhiên.

6.2.3 Ma trận xác suất chuyển bậc cao, Phương trình Chapman–Kolmogorov

Đặt: $p_{ij}^{(k)} = P\left\{ X_{n+k} = j \middle| X_n = i \right\} = P\left\{ X_k = j \middle| X_0 = i \right\}$. (6.16)

Đó là xác suất sau k bước hệ sẽ chuyển từ trạng thái i sang trạng thái j.

Định nghĩa 6.3: *Ma trận vuông* $P^{(k)} = \left[p_{ij}^{(k)} \right]$ *gọi là ma trận xác suất chuyển sau* k *bước.*

Ký hiệu $P^{(0)} = I,\ P^{(1)} = P,\ I$ là ma trận đơn vị (có các phần tử trên đường chéo bằng 1 và các phần tử ở vị trí còn lại bằng 0).

Tương tự ma trận xác suất chuyển P, số hàng số cột của $P^{(k)}$ có thể vô hạn nếu không gian trạng thái E có vô số đếm được các phần tử. Nếu không gian trạng thái E hữu hạn thì ma trận xác suất chuyển sau k bước $P^{(k)}$ cũng là ma trận Markov (xem bài tập 6.8).

Định lý 6.1: Với mọi $n \geq 0$, ta có:

$$P^{(n+1)} = PP^{(n)} = P^{(n)}P \tag{6.17}$$

Từ đó suy ra:

$$P^{(n)} = P^n \tag{6.18}$$

Chứng minh: Áp dụng công thức xác xuất đầy đủ (1.19) ta có

Với $\forall$i, j: $p_{ij}^{(n+1)} = P\{X_{n+1} = j | X_0 = i\}$

$$= \sum_{k \in E} P\{X_{n+1} = j | X_0 = i, X_1 = k\} P\{X_1 = k | X_0 = i\}$$

$$= \sum_{k \in E} P\{X_{n+1} = j | X_1 = k\} P\{X_1 = k | X_0 = i\}$$

$$= \sum_{k \in E} p_{ik} p_{kj}^{(n)} \Rightarrow P^{(n+1)} = PP^{(n)}.$$

Ta cũng có: $p_{ij}^{(n+1)} = P\{X_{n+1} = j | X_0 = i\}$

$$= \sum_{k \in E} P\{X_{n+1} = j | X_0 = i, X_n = k\} P\{X_n = k | X_0 = i\}$$

$$= \sum_{k \in E} P\{X_{n+1} = j | X_n = k\} P\{X_n = k | X_0 = i\}$$

$$= \sum_{k \in E} p_{ik}^{(n)} p_{kj} \Rightarrow P^{(n+1)} = P^{(n)}P.$$

Từ (6.17) suy ra $P^{(2)} = PP = P^2$, bằng quy nạp ta có $P^{(n)} = P^n$ với mọi $n = 0,1,2,\ldots$.

Từ công thức (6.18) và đẳng thức:

$$P^{n+m} = P^n P^m, \ \forall n, m \geq 0$$

ta có thể viết các phần tử tương ứng dưới dạng

$$p_{ij}^{(n+m)} = \sum_k p_{ik}^{(n)} p_{kj}^{(m)} \tag{6.19}$$

Công thức (6.19) được gọi là Phương trình Chapman-Kolmogorov.

Phương trình Chapman-Kolmogorov giải thích quy luật chuyển trạng thái của chuỗi Markov như sau: hệ chuyển từ trạng thái i sang trạng thái j sau $n+m$ bước có thể đạt được bằng cách chuyển từ trạng thái i sang trạng thái trung gian k trong n bước (với xác suất $p_{ik}^{(n)}$) và tiếp tục chuyển từ trạng thái k sang trạng thái j trong m bước (với xác suất $p_{kj}^{(m)}$). Hơn nửa biến cố "chuyển từ trạng thái i sang trạng thái trung gian k trong n bước" và biến cố "chuyển từ trạng thái k sang trạng thái j trong m bước" là độc lập, vì vậy xác suất chuyển từ i sang j sau $n+m$ bước qua i,k,j bằng $p_{ik}^{(n)}p_{kj}^{(m)}$. Cuối cùng xác suất chuyển từ i sang j có được bằng cách lấy tổng theo k, k chạy trong không gian các trạng thái của chuỗi.

6.2.4 Phân bố xác suất của chuỗi tại thời điểm thứ *n*

Giả sử không gian trạng thái có dạng $E=\{0,1,2,...\}$, tại thời điểm n ta có hàm khối lượng xác suất của biến ngẫu nhiên X_n theo công thức (2.8) $p_{X_n}(k)=P\{X_n=k\}$. Tuy nhiên để dễ tính toán ta ký hiệu các khối lượng xác suất dưới dạng ma trận.

Ma trận hàng

$$\mathbf{P}(n)=\begin{bmatrix}p_0(n) & p_1(n) & p_2(n) & \cdots\end{bmatrix}, \tag{6.20a}$$

với $$p_j(n)=P\{X_n=j\},\ n=0,1,2,\ldots. \tag{6.20b}$$

gọi là ma trận phân bố của hệ tại thời điểm n hoặc phân bố của X_n.

Các phần tử của ma trận hàng $\mathrm{P}(n)$ thỏa mãn điều kiện

$$p_k(n)\ge 0;\ \sum_k p_k(n)=1$$

Khi $n=0$, $\mathbf{P}(0)=\begin{bmatrix}p_0(0) & p_1(0) & p_2(0) & \cdots\end{bmatrix}$ được gọi là ma trận phân bố ban đầu.

Định lý sau đây chỉ ra rằng ma trận phân bố của hệ tại thời điểm bất kỳ có thể tính được thông qua ma trận phân bố đầu và ma trận xác suất chuyển.

Định lý 6.2: Với mọi $n \geq 0$, $m \geq 0$:

$$\mathbf{P}(n) = \mathbf{P}(0)P^{(n)} \tag{6.21}$$

$$\mathbf{P}(n+1) = \mathbf{P}(n)P \tag{6.22}$$

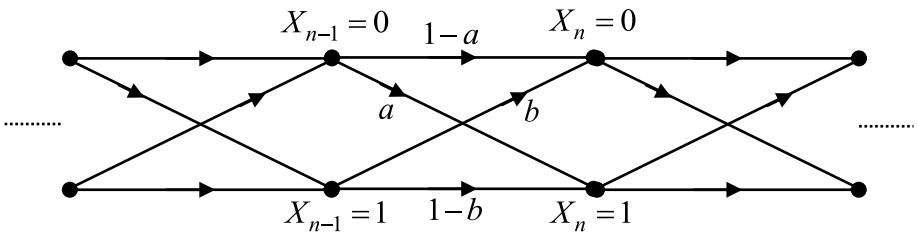
$$\mathbf{P}(n+m) = \mathbf{P}(n)P^{(m)}. \tag{6.23}$$

Chứng minh: Từ định lý 6.1 ta suy ra 3 điều trên là tương đương. Vì vậy để chứng minh định lý 6.2 ta chỉ cần chứng minh (6.23).

$$p_j(n+m) = P\{X_{n+m} = j\}$$
$$= \sum_{i \in E} P\{X_n = i\} P\{X_{n+m} = j \,|\, X_n = i\} = \sum_{i \in E} p_i(n) p_{ij}^{(m)}$$

Vậy chuỗi Markov rời rạc thuần nhất hoàn toàn được xác định bởi ma trận xác suất chuyển một bước P và ma trận phân bố ban đầu $\mathbf{P}(0)$.

Ví dụ 6.2: Một mạng viễn thông gồm một dãy các trạm chuyển tiếp các kênh viễn thông nhị phân cho trong sơ đồ sau:



Hình 6.3: Mạng viễn thông nhị phân

Trong đó X_n ký hiệu mã số nhị phân đầu ra của trạm thứ n và X_0 ký hiệu mã số nhị phân đầu vào của trạm đầu tiên.

Đây là 1 mô hình chuỗi Markov có không gian trạng thái $E = \{0,1\}$, tập chỉ số $T = \{0,1,...,n,...\}$. Ma trận xác suất chuyển của mạng viễn thông này thường được gọi là ma trận kênh:

$$P=\begin{bmatrix}1-a & a\\ b & 1-b\end{bmatrix};\ 0<a<1,\ 0<b<1.$$

Giả sử $a=0,1$, $b=0,2$ và phân bố xác suất đầu $P\{X_0=0\}=P\{X_0=1\}=0,5$ (hai tín hiệu 0, 1 đồng khả năng).

a) Tìm ma trận xác suất chuyển sau 2 bước;

b) Tìm phân bố của trạm thứ hai X_2.

Giải:

a) $P^{(2)}=\begin{bmatrix}0,9 & 0,1\\ 0,2 & 0,8\end{bmatrix}\begin{bmatrix}0,9 & 0,1\\ 0,2 & 0,8\end{bmatrix}=\begin{bmatrix}0,83 & 0,17\\ 0,34 & 0,66\end{bmatrix}.$

b) $\mathbf{P}(2)=\mathbf{P}(0)P^{(2)}=\begin{bmatrix}0,5 & 0,5\end{bmatrix}\begin{bmatrix}0,83 & 0,17\\ 0,34 & 0,66\end{bmatrix}=\begin{bmatrix}0,585 & 0,415\end{bmatrix}.$

Như vậy có 58,5% tín hiệu 0 và 41,5% tín hiệu 1 ở đầu ra của trạm thứ hai, mặc dù đầu vào ở trạm đầu tiên hai tín hiệu này xuất hiện đồng khả năng.

Ví dụ 6.3: (Mô hình hòa nhập cộng đồng của các bệnh nhân tâm thần được xuất viện).

Các chuyên gia y tế thường tránh chuyển các bệnh nhân tâm thần lâu năm được xuất viện trực tiếp từ bệnh viện đến với cộng đồng. Chẳng hạn ở Billings, Montana, người ta thực hiện như sau: Trước hết người ta chuyển bệnh nhân đến ở tại khu vực được chăm sóc 24/24 giờ. Nếu tình trạng sức khỏe của bệnh nhân tiến triển tốt đáp ứng những tiêu chí đòi hỏi thì được chuyển đến nhóm 40 giờ, tức là được chăm sóc 5 ngày trong 1 tuần và 1 ngày 8 giờ. Nếu tình trạng bệnh nhân tiếp tục tiến triển tốt hơn thì sẽ được đưa đến nhóm có sự tương tác giao tiếp cao hơn, ở đây bệnh nhân được luyện tập tự chủ hành vi của mình. Cuối cùng khi được coi là khỏi bệnh hoàn toàn thì được đưa về hòa nhập với cộng đồng.

Drachman (1981) đã phân tích các dữ liệu thu thập được ở Billings từ 1/1/1978 đến 31/5/1979 và nhận thấy rằng dữ liệu tuân theo mô hình chuỗi Markov với ma trận xác suất chuyển như sau:

$$P = \begin{array}{c} \\ H \\ I \\ 24 \\ 40 \\ A \\ C \end{array} \begin{array}{c} \begin{array}{cccccc} H & I & 24 & 40 & A & C \end{array} \\ \begin{bmatrix} 0,7143 & 0,0714 & 0,0714 & 0,0000 & 0,0000 & 0,1429 \\ 0,1177 & 0,0588 & 0,2941 & 0,1177 & 0,0000 & 0,4118 \\ 0,0109 & 0,0109 & 0,7283 & 0,0652 & 0,0000 & 0,1848 \\ 0,0213 & 0,0213 & 0,0213 & 0,7660 & 0,0426 & 0,1277 \\ 0,0000 & 0,0172 & 0,0172 & 0,0172 & 0,7931 & 0,1552 \\ 0,0136 & 0,0442 & 0,0578 & 0,0034 & 0,0272 & 0,8537 \end{bmatrix} \end{array}$$

trong đó các trạng thái H (ở bệnh viện), I (bắt đầu chuyển khỏi bệnh viện), 24 (nhóm chăm sóc 24/24), 40 (nhóm chăm sóc 40 giờ/1 tuần), A (nhóm được tương tác giao tiếp) và C (nhóm được đưa về cộng đồng). Ở đây 12 tuần được qui tròn thành 3 tháng.

Áp dụng công thức (6.18) ta tính được:

$$P^6 = \begin{array}{c} \\ H \\ I \\ 24 \\ 40 \\ A \\ C \end{array} \begin{array}{c} \begin{array}{cccccc} H & I & 24 & 40 & A & C \end{array} \\ \begin{bmatrix} 0,1723 & 0,0424 & 0,2002 & 0,0585 & 0,0372 & 0,4894 \\ 0,0678 & 0,0359 & 0,2032 & 0,1010 & 0,0600 & 0,5323 \\ 0,0454 & 0,0323 & 0,2539 & 0,1167 & 0,0507 & 0,5010 \\ 0,0548 & 0,0330 & 0,1256 & 0,2373 & 0,1046 & 0,4447 \\ 0,0282 & 0,0313 & 0,1180 & 0,0592 & 0,2870 & 0,4762 \\ 0,0489 & 0,0374 & 0,1758 & 0,0548 & 0,0758 & 0,6073 \end{bmatrix} \end{array}.$$

Dữ liệu ban đầu $O_1 = \begin{bmatrix} 1 & 0 & 15 & 8 & 10 & 53 \end{bmatrix}$ (biểu diễn theo tần số), áp dụng công thức (6.17) tính được:

$$e_7 = O_1 P^6 = \begin{bmatrix} 4,17 & 3,09 & 15,51 & 7,20 & 8,52 & 48,51 \end{bmatrix},$$

ở đây 17 tháng được làm tròn thành 6 chu kỳ 12 tuần.

Giá trị thức tế $O_7 = \begin{bmatrix} 5 & 1 & 15 & 7 & 8 & 51 \end{bmatrix}$.

Sử dụng phép thử "khi bình phương" để so sánh, người ta thấy rằng giá trị lý thuyết e_7 phù hợp với giá trị thức tế O_7. Điều này chứng tỏ chuỗi Markov phù hợp với mô hình trên.

6.2.5 Một số mô hình chuỗi Markov quan trọng

6.2.5.1 Mô hình phục vụ đám đông

Xét mô hình phục vụ đám đông (lý thuyết sắp hàng). Khách đến sắp hàng chờ phục vụ theo nguyên tắc FIFO (first in first out) và trong mỗi chu kỳ cửa hàng chỉ phục vụ một khách. Số khách đến trong chu kỳ thứ n là biến ngẫu nhiên ξ_n. Giả sử ξ_1, ξ_2,... là các biến ngẫu nhiên độc lập cùng phân bố với biến ngẫu nhiên ξ có phân bố xác suất.

$$P\{\xi = k\} = a_k\ ;\ k = 0,1,2,...\ ;\ \ a_k > 0;\ \sum_k a_k = 1\,. \qquad (6.24)$$

Trạng thái của hệ (cửa hàng) là số khách xếp hàng chờ phục vụ tại thời điểm đầu của mỗi chu kỳ (khi một khách hàng vừa được phục vụ xong). Nếu hiện tại hệ ở trạng thái i và sau 1 chu kỳ hệ rơi vào trạng thái j thì:

$$j = \begin{cases} i - 1 + \xi & \text{nếu } \ i \geq 1 \\ \xi & \text{nếu } \ i = 0. \end{cases} \qquad (6.25)$$

Ký hiệu X_n là số khách hàng tại thời điểm đầu của chu kỳ thứ n thì:

$$X_{n+1} = \left(X_n - 1\right)^+ + \xi_n\,,\ \text{trong đó}\ \ X^+ = \max(0, X)\,,$$

Từ (6.24), (6.25) suy ra

$$P\left\{X_{n+1} = j \middle| X_n = i\right\} = \begin{cases} P\{\xi_n = j + 1 - i\} & \text{nếu } i > 0 \\ P\{\xi_n = j\} & \text{nếu } i = 0 \end{cases}$$

$$= \begin{cases} a_j & \text{nếu } \ i = 0, j \geq 0 \\ 0 & \text{nếu } \ j + 1 < i \\ a_{j+1-i} & \text{nếu } \ j + 1 \geq i > 0 \end{cases}$$

Vì các quá trình đến ξ_n độc lập do đó xác suất chuyển $p_{ij} = P\left\{X_{n+1} = j \middle| X_n = i\right\}$ thỏa mãn điều kiện (6.7), hơn nữa các biến ngẫu nhiên ξ_n có cùng phân bố do đó xác suất chuyển p_{ij} thuần nhất theo thời gian.

Vậy $\{X_n; n = 0,1,...\}$ là chuỗi Markov thuần nhất với ma trận xác suất chuyển:

$$P = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{6.26}$$

6.2.5.2 Mô hình kiểm kê (Inventory Model)

Giả thiết phải dự trữ trong kho một loại hàng nào đó để đáp ứng nhu cầu liên tục của khách hàng. Hàng được nhập kho tại cuối mỗi chu kỳ $n = 0,1,2,...$ Giả sử tổng số lượng hàng cần phải đáp ứng nhu cầu trong chu kỳ n là biến ngẫu nhiên ξ_n có phân bố độc lập với chu kỳ thời gian, nghĩa là dãy biến ngẫu nhiên $\{\xi_n\}$ độc lập có cùng phân bố với ξ.

$$P\{\xi = k\} = a_k \ ; a_k > 0 \text{ và } \sum_k a_k = 1. \tag{6.27}$$

Mức hàng dự trữ được kiểm kê tại cuối mỗi chu kỳ. Cách nhập hàng căn cứ vào 2 chỉ số tiêu chuẩn s và S $(s < S)$ như sau: Nếu ở cuối mỗi chu kỳ lượng hàng dự trữ $\le s$ thì ngay tức khắc nhập hàng để có số hàng dự trữ bằng S; Nếu hàng hiện có $> s$ thì không cần nhập hàng. Giả sử số nhu cầu trong mỗi chu kỳ không vượt quá S.

Ký hiệu X_n là lượng hàng hiện có tại cuối chu kỳ n và trước khi nhập hàng, như vậy:

$$X_{n+1} = \begin{cases} X_n - \xi_{n+1} & \text{nếu} \quad s < X_n \le S \\ S - \xi_{n+1} & \text{nếu} \quad X_n \le s. \end{cases} \tag{6.28}$$

Các trạng thái của quá trình $\{X_n, n \ge 0\}$ là các số lượng hàng dự trữ:

$$S, S-1, ..., 1, 0, -1, -2, ...$$

trong đó giá trị âm là nhu cầu chưa được phục vụ mà sẽ được đáp ứng ngay sau khi nhập hàng:

$$p_{ij} = P\{X_{n+1} = j \,|\, X_n = i\} = \begin{cases} P\{\xi_{n+1} = i - j\} & \text{nếu } s < i \le S \\ P\{\xi_{n+1} = S - j\} & \text{nếu } i \le s. \end{cases} \quad (6.29)$$

Ví dụ 6.4: Xét mô hình kiểm kê phụ tùng thay thế, trong đó yêu cầu có thể là 0, 1 hoặc 2 đơn vị phụ tùng cần thay thế trong một chu kỳ bất kỳ với phân bố xác suất như sau:

$$P\{\xi = 0\} = 0,5;\ P\{\xi = 1\} = 0,4;\ P\{\xi = 2\} = 0,1$$

và giả sử $s = 0; S = 2$.

Không gian trạng thái sẽ là $E = \{-1, 0, 1, 2\}$.

Ta có: $p_{ij} = P\{X_{n+1} = j \,|\, X_n = i\} = \begin{cases} P\{\xi = i - j\} & \text{nếu } 0 < i \le 2 \\ P\{\xi = 2 - j\} & \text{nếu } i \le 0. \end{cases}$

$$p_{-1,-1} = P\{X_{n+1} = -1 \,|\, X_n = -1\} = P(\varnothing) = 0,$$

$$p_{-1,0} = P\{X_{n+1} = 0 \,|\, X_n = -1\} = P(\xi = 2) = 0,1,$$

$$p_{-1,1} = P\{X_{n+1} = 1 \,|\, X_n = -1\} = P(\xi = 1) = 0,4,$$

. . . . . .

Ma trận xác suất chuyển:

$$P = \begin{bmatrix} 0,0 & 0,1 & 0,4 & 0,5 \\ 0,0 & 0,1 & 0,4 & 0,5 \\ 0,1 & 0,4 & 0,5 & 0,0 \\ 0,0 & 0,1 & 0,4 & 0,5 \end{bmatrix}$$

Mô hình di động ngẫu nhiên được xét trong mục 6.4.

6.2.6 Phân bố dừng, phân bố giới hạn, phân bố ergodic

Định nghĩa 6.4: *Ma trận hàng* $\mathbf{P}^* = [p_1 \quad p_2 \quad ...]$ *được gọi là phân bố dừng của chuỗi Markov với ma trận xác suất chuyển* P *nếu thoả mãn 2 điều kiện:*

$$\begin{cases} \mathbf{P}^* = \mathbf{P}^* P & (a) \\ p_j \geq 0, \sum_j p_j = 1 & (b) \end{cases} \tag{6.30}$$

Từ (6.30 a) suy ra

$$\mathbf{P}^* = \mathbf{P}^* P = \mathbf{P}^* P^2 = ... = \mathbf{P}^* P^n \ ; \forall \ n \, .$$

Do đó nếu lấy $\mathbf{P}^*$ làm phân bố đầu của chuỗi Markov thì $\mathbf{P}^*(n) = \mathbf{P}^*, \forall n$. Như vậy nếu chuỗi Markov có phân bố dừng tại thời điểm n_0 nào đó thì hệ sẽ có phân bố xác suất ổn định sau mọi bước chuyển kể từ thời điểm n_0.

Điều kiện (6.30-a) có thể viết lại dưới dạng

$$P^t \, \mathbf{P}^{*t} = \mathbf{P}^{*t} \tag{6.31}$$

trong đó ma trận cột $\mathbf{P}^{*t}$ là ma trận chuyển vị của $\mathbf{P}^*$.

Công thức (6.31) cho thấy phân bố dừng $\mathbf{P}^*$ là véc tơ riêng với giá trị riêng bằng 1 của ma trận P^t.

Định nghĩa 6.5: *Ta nói rằng chuỗi Markov với ma trận xác suất chuyển P có phân bố giới hạn là* $[p_1 \quad p_2 \quad ...]$ *nếu thoả mãn 2 điều kiện*:

1) Với mọi j tồn tại giới hạn $\lim_{n\to\infty} p_{ij}^{(n)} = p_j$ không phụ thuộc i,

(6.32)

2) $\sum_j p_j = 1$, $p_j \geq 0$, (6.33)

Nếu điều kiện (6.33) được thay bởi

3) $\sum_j p_j = 1$, $p_j > 0$ (6.34)

thì chuỗi Markov được gọi là *có tính ergodic còn* $[p_1 \quad p_2 \quad \ldots]$ *là phân bố ergodic*.

Nhận xét 6.1:

- Nếu phân bố của X_{n_0} (ở thời điểm thứ n_0) của chuỗi là phân bố dừng thì từ thời điểm này trở đi phân bố của chuỗi không thay đổi; nghĩa là với mọi $m \geq n_0$, X_m và X_{n_0} có cùng phân bố.

- Phân bố giới hạn là phân bố hệ sẽ đạt được khi thời gian tiến đến vô cùng. Phân bố giới hạn chỉ phụ thuộc ma trận xác suất chuyển, không phụ thuộc phân bố đầu (ví dụ 6.6). Trong thực tế có thể đến thời điểm nào đó trở đi chuỗi sẽ đạt được phân bố giới hạn. Ví dụ 6.5 sau đây chứng tỏ với $n = 20$ thì chuỗi đạt được phân bố giới hạn.

- Phân bố ergodic là phân bố giới hạn với xác suất dương tại mọi trạng thái của chuỗi.

Ví dụ 6.5: Có 3 mạng điện thoại di động A, B, C cùng khai thác thị trường. Tỉ lệ chiếm lĩnh thị trường hiện tại tương ứng là 40%, 30% và 30%. Theo thống kê người ta thấy xác suất thay đổi mạng của khách hàng trong mỗi tháng như sau:

$$P = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{bmatrix} 0,6 & 0,3 & 0,1 \\ 0,1 & 0,8 & 0,1 \\ 0,1 & 0,2 & 0,7 \end{bmatrix} \end{array}$$

Áp dụng công thức (6.18) và (6.21) ta tính được phân bố tại thời điểm thứ n:

$$\mathbf{P}(n) = \mathbf{P}(0)P^{(n)}$$

trong các trường hợp sau.

$n = 0$

$$\mathbf{P}(0) = \begin{bmatrix} 0,4 & 0,3 & 0,3 \end{bmatrix},$$

$n = 1$

$$P = \begin{bmatrix} 0,6 & 0,3 & 0,1 \\ 0,1 & 0,8 & 0,1 \\ 0,1 & 0,2 & 0,7 \end{bmatrix}$$

$$\mathbf{P}(1) = \mathbf{P}(0)P = \begin{bmatrix} 0,35 & 0,43 & 0,22 \end{bmatrix},$$

$n = 6$

$$P^6 = \begin{bmatrix} 0,2125 & 0,5492 & 0,2383 \\ 0,1969 & 0,5648 & 0,2383 \\ 0,1969 & 0,5181 & 0,2853 \end{bmatrix}$$

$$\mathbf{P}(6) = \mathbf{P}(0)P^6 = \begin{bmatrix} 0,2047 & 0,5476 & 0,2477 \end{bmatrix}$$

$n = 12$

$$P^{12} = \begin{bmatrix} 0,2002 & 0,5503 & 0,2495 \\ 0,2000 & 0,5506 & 0,2495 \\ 0,2000 & 0,5484 & 0,2516 \end{bmatrix}$$

$$\mathbf{P}(12) = \mathbf{P}(0)P^{12} = \begin{bmatrix} 0,2001 & 0,550 & 0,2499 \end{bmatrix}$$

$n = 18$

$$P^{18} = \begin{bmatrix} 0,2000 & 0,5500 & 0,2500 \\ 0,2000 & 0,5500 & 0,2500 \\ 0,2000 & 0,5499 & 0,2501 \end{bmatrix}$$

$$\mathbf{P}(18) = \mathbf{P}(0)P^{18} = \begin{bmatrix} 0,2000 & 0,550 & 0,2500 \end{bmatrix}$$

$n = 20$

$$P^{20} = \begin{bmatrix} 0,2000 & 0,5500 & 0,2500 \\ 0,2000 & 0,5500 & 0,2500 \\ 0,2000 & 0,5500 & 0,2500 \end{bmatrix}$$

$$\mathbf{P}(20) = \mathbf{P}(0)P^{20} = \begin{bmatrix} 0,20 & 0,55 & 0,25 \end{bmatrix}.$$

Ta thấy rằng khi n càng lớn xác suất trên mỗi cột càng gần bằng nhau và đạt được phân bố giới hạn khi $n = 20$.

Vậy thị trường đạt trạng thái ổn định với tỉ lệ tương ứng 20%, 55% và 25%. Phân bố giới hạn này chỉ phụ thuộc ma trận xác suất chuyển và không phụ thuộc phân bố ban đầu.

Ví dụ sau đây cũng minh họa thêm về điều đó.

Ví dụ 6.6: Về sự bình đẳng trong giáo dục giữa các nhóm chủng tộc.

Trên cơ sở báo cáo điều tra dân số của Văn phòng Điều tra dân số Hoa Kỳ năm 1960, hai tác giả Lieberson và Fuguitt (1967) đã xác định được ma trận chuyển trình độ học vấn giữa hai thế hệ khi so sánh tình trạng học vấn của nhóm thanh niên độ tuổi 20-24 với trình độ học vấn của bố của họ:

$$P = \begin{array}{c} \\ \text{Dưới ĐH} \\ \text{ĐH} \\ \text{Trên ĐH} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Dưới ĐH} & \text{ĐH} & \text{Trên ĐH} \end{array} \\ \begin{bmatrix} 0{,}43 & 0{,}34 & 0{,}23 \\ 0{,}10 & 0{,}36 & 0{,}54 \\ 0{,}05 & 0{,}15 & 0{,}80 \end{bmatrix} \end{array}$$

Hai tác giả đồng ý rằng có hai loại bất lợi đối với các nhóm chủng tộc và dân tộc. Loại bất lợi thứ nhất bắt nguồn từ nguồn gốc chủng tộc và dân tộc mà kết quả là có sự khác nhau giữa ma trận chuyển của nhóm người da trắng và nhóm người da màu. Ngay cả khi sự phân biệt chủng tộc bị loại bỏ thì vẫn còn loại bất lợi thứ hai đó là vị trí xã hội và thu nhập của người da màu thấp hơn nhiều so với người da trắng. Nói cách khác ngay cả khi ma trận chuyển về học vấn giữa hai thế hệ P (ma trận xác suất chuyển) được xem là như nhau giữa hai nhóm thì điều kiện ban đầu $\mathbf{P}(0)$ (phân bố đầu) cũng khác nhau.

Bảng kết quả sau được giả định rằng ma trận chuyển P giữa hai nhóm da trắng và da màu là như nhau nhưng có xuất phát điểm khác nhau. Chỉ số khác nhau trong bảng là tỷ lệ % khoảng cách mà hai nhóm cần phải thay đổi để đạt được phân bố trình độ học vấn bằng nhau.

		% Dưới ĐH	% ĐH	% Trên ĐH	Chỉ số % khác nhau
P(1) (1960)	Da trắng	46	31	23	29
	Da màu	75	16	09	
P(2)	Da trắng	24	30	46	13
	Da màu	34	33	33	
P(3)	Da trắng	16	26	58	6
	Da màu	20	28	52	
P(4)	Da trắng	12	23	64	3
	Da màu	14	25	61	
P(5)	Da trắng	11	22	68	1
	Da màu	11	23	66	
P(6)	Da trắng	10	22	68	1
	Da màu	11	22	67	
P(7)	Da trắng	10	21	69	1
	Da màu	10	22	68	
P(8)	Da trắng	10	21	69	0
	Da màu	10	21	69	

Như vậy không phụ thuộc vào xuất phát điểm, sau 8 thế hệ các nhóm người trong cộng đồng đều có trình độ học vấn như nhau theo tỷ lệ 10% dưới ĐH, 21% ĐH và 69% trên ĐH.

Định lý 6.3: Nếu tồn tại phân bố giới hạn thì đó là phân bố dừng duy nhất.

Chứng minh: Giả sử $[p_1 \quad p_2 \quad \dots]$ là phân bố giới hạn, khi đó với mọi j ta có:

$$p_j = \lim_{n\to\infty} p_{ij}^{(n+1)} = \lim_{n\to\infty}\left(\sum_k p_{ik}^{(n)} p_{kj}\right) = \sum_k p_k p_{kj}$$

$\Rightarrow [p_1 \quad p_2 \quad \dots] = [p_1 \quad p_2 \quad \dots]P$. Do đó $[p_1 \quad p_2 \quad \dots]$ là một phân bố dừng.

Ngược lại giả sử $\left[\overline{p}_1 \quad \overline{p}_2 \quad \dots\right]$ là một phân bố dừng bất kỳ của chuỗi Markov này thì:

$$\overline{p}_j = \sum_k \overline{p}_k p_{kj} = \sum_k \overline{p}_k p_{kj}^{(2)} = \dots. = \sum_k \overline{p}_k p_{kj}^{(n)}$$

$$\Rightarrow \quad \overline{p}_j = \lim_{n\to\infty}\left(\sum_k \overline{p}_k p_{kj}^{(n)}\right) = \sum_k \overline{p}_k p_j = \overline{p}_j$$

Nghĩa là phân bố giới hạn là phân bố dừng duy nhất.

Định lý sau đây xác định điều kiện cần và đủ để một chuỗi Markov với hữu hạn trạng thái tồn tại phân bố ergodic.

Định lý 6.4: Nếu chuỗi Markov có không gian trạng thái hữu hạn thì chuỗi này là ergodic khi và chỉ khi tồn tại n_0 sao cho $\min_{i,j} p_{ij}^{(n_0)} > 0$.

Nhận xét 6.2: Từ định lý 6.3 và 6.4 ta thấy rằng nếu chuỗi Markov có ma trận xác suất chuyển $P = \left[p_{ij}\right]$ thỏa mãn điều kiện tồn tại n_0 sao cho $\min_{i,j} p_{ij}^{(n_0)} > 0$ thì chuỗi này là ergodic. Phân bố ergodic cũng là phân bố dừng duy nhất, đó là nghiệm của hệ phương trình:

$$\begin{cases} [x_1 \quad x_2 \quad ...] = [x_1 \quad x_2 \quad ...]P \\ x_j > 0, \ \sum_j x_j = 1. \end{cases}$$

$$\text{hay} \quad \begin{cases} P^t \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} \\ x_j > 0, \ \sum_j x_j = 1. \end{cases} \tag{6.35}$$

Giải hệ phương trình (6.35) cho trường hợp ví dụ 6.6 ta cũng thu được phân bố dừng tương ứng $\mathbf{P}^* = [0,20 \quad 0,550 \quad 0,25]$.

Ví dụ 6.7: Xét chuỗi Markov ở ví dụ 6.2, ma trận xác suất chuyển:

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}, \ 0 < a, b < 1$$

Theo định lý 6.4 chuỗi Markov có tính ergodic với phân bố ergodic là nghiệm của hệ phương trình:

$$\begin{cases} \begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ x_1 + x_2 = 1 \end{cases} \Leftrightarrow \begin{cases} -ax_1 + bx_2 = 0 \\ x_1 + x_2 = 1 \end{cases} \Leftrightarrow \begin{cases} x_1 = \dfrac{b}{a+b} \\ x_2 = \dfrac{a}{a+b} \end{cases}$$

Mặt khác cũng có thể tìm phân bố ergodic bằng cách tính trực tiếp ma trận chuyển sau n bước:

$$P^n = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}^n = \frac{1}{a+b} \left\{ \begin{bmatrix} b & a \\ b & a \end{bmatrix} + (1-a-b)^n \begin{bmatrix} a & -a \\ -b & b \end{bmatrix} \right\} \tag{6.36}$$

$$\Rightarrow \lim_{n \to \infty} P^{(n)} = \begin{bmatrix} \dfrac{b}{a+b} & \dfrac{a}{a+b} \\ \dfrac{b}{a+b} & \dfrac{a}{a+b} \end{bmatrix}$$

Do đó chuỗi tồn tại phân bố giới hạn.

Để chứng minh (6.36) ta có thể tính theo một trong các cách sau:

- Quy nạp theo n.

- Sử dụng công thức: nếu $AB = BA$ thì $(A+B)^n = \sum_{k=0}^{n} C_n^k A^k B^{n-k}$

và bằng cách đặt

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix} = \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix};$$

$$A = \begin{bmatrix} -a & a \\ b & -b \end{bmatrix} \Rightarrow A^k = (-a-b)^{k-1} A$$

$$P^n = \sum_{k=0}^{n} C_n^k A^k = I + \sum_{k=1}^{n} C_n^k A^k = I + \left(\sum_{k=1}^{n} C_n^k (-a-b)^{k-1} \right) A$$

$$= I + \frac{1}{-(a+b)} \left((1-a-b)^n - 1 \right) A$$

$$= \frac{1}{a+b} \left\{ \begin{bmatrix} b & a \\ b & a \end{bmatrix} + (1-a-b)^n \begin{bmatrix} a & -a \\ -b & b \end{bmatrix} \right\}$$

Ví dụ 6.8: Trong một bài báo viết năm 1913 A. A. Markov đã chọn 1 dãy gồm 20.000 chữ cái trong trường ca Evghenhi Onheghin của A. X. Puskin và thấy rằng các chữ cái này chuyển đổi liên tiếp theo hai trạng thái nguyên âm (Na) và phụ âm (Pa) với ma trận xác suất chuyển là:

$$P = \begin{matrix} Na \\ Pa \end{matrix} \begin{bmatrix} 0,128 & 0,872 \\ 0,663 & 0,337 \end{bmatrix}$$
$$\begin{matrix} Na & Pa \end{matrix}$$

Phân bố giới hạn (cũng là phân bố dừng) của chuỗi Markov này là:

$$P(Na) = \frac{0,663}{0,872 + 0,663} = 0,423$$

$$P(Pa) = \frac{0,872}{0,872 + 0,663} = 0,568$$

Vậy có khoảng 42,3% nguyên âm và 56,8% phụ âm trong tác phẩm trên.

6.3 PHÂN LOẠI TRẠNG THÁI CHUỖI MARKOV

Định lý 6.4 cho ta dấu hiệu nhận biết một chuỗi Markov hữu hạn trạng thái tồn tại phân bố ergodic. Trong trường hợp tổng quát, bằng cách phân tích trạng thái của chuỗi Markov ta sẽ tìm điều kiện để chuỗi tồn tại phân bố dừng, phân bố giới hạn hoặc phân bố ergodic thỏa mãn các điều kiện (6.31), (6.34).

6.3.1 Các trạng thái liên thông và sự phân lớp

Định nghĩa 6.6: *Ta nói rằng trạng thái j đạt được (accessible) từ trạng thái i nếu tồn tại $n \geq 0$ sao cho $p_{ij}^{(n)} > 0$ (xác suất của biến cố sau n bước hệ chuyển từ trạng thái i sang trạng thái j nhận giá trị dương). Ký hiệu $i \to j$.*

Quy ước $p_{ii}^{(0)} = 1$ và $p_{ij}^{(0)} = 0$ khi $i \neq j$.

Hai trạng thái i và j được gọi là liên thông (communicate) với nhau nếu $i \to j$ và $j \to i$, lúc đó ta ký hiệu $i \leftrightarrow j$.

Có thể chứng minh được rằng $\leftrightarrow$ là một quan hệ tương đương (có tính phản xạ, đối xứng và bắc cầu) trên tập các trạng thái. Do đó ta có thể phân hoạch không gian trạng thái thành các lớp tương đương. Các lớp tương đương này rời nhau, hai trạng thái bất kỳ cùng một lớp thì liên thông với nhau, còn hai trạng thái thuộc hai lớp khác nhau không thể liên thông với nhau.

Định nghĩa 6.7: *Chuỗi Markov được gọi là tối giản (irreducible) nếu hai trạng thái bất kỳ của không gian trạng thái liên thông với nhau.*

Như vậy chuỗi Markov tối giản chỉ có một lớp tương đương.

Ví dụ 6.9: Cho chuỗi Markov với ma trận xác suất chuyển

$$P=\begin{bmatrix} 1/3 & 2/3 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}=\begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix}$$

Không gian trạng thái $E=\{1,2,3,4,5\}$ phân thành hai lớp $E_1=\{1,2\}$, $E_2=\{3,4,5\}$. Có thể xem E_1,E_2 lần lượt là hai không gian trạng thái của hai chuỗi Markov với ma trận xác suất chuyển tương ứng là P_1 và P_2, hai chuỗi Markov này tối giản.

Một cách tổng quát, giả sử không gian trạng thái được tách thành các lớp tương đương

$$E=E_1\cup E_2\cup\cdots$$

Khi đó có thể xem mỗi E_k $(k=1,2,...)$ là không gian trạng thái của chuỗi Markov tối giản với ma trận xác suất chuyển là ma trận con P_k ứng với phép phân hoạch tương ứng của không gian trạng thái. Vì thế $E_1,E_2,...$ được gọi là các lớp tối giản của chuỗi.

6.3.2 Trạng thái tuần hoàn và không tuần hoàn

Định nghĩa 6.8: *Ta định nghĩa chu kỳ của trạng thái i là*:

$$d(i)=UCLN\left\{n\geq 1: p_{ii}^{(n)}>0\right\} \tag{6.36}$$

trong đó UCLN viết tắt của "ước chung lớn nhất".

Nếu $p_{ii}^{(n)}=0$ đối với mọi $n\geq 1$ thì đặt $d(i)=0$.

Nếu $d(i)>1$ thì trạng thái i được gọi là *trạng thái tuần hoàn* với chu kỳ $d(i)$.

Nếu $d(i)=1$ thì trạng thái i được gọi là *trạng thái không tuần hoàn*. Rõ ràng nếu $p_{ii}>0$ thì i là trạng thái không tuần hoàn.

6.3.3 Biểu đồ chuyển trạng thái của chuỗi Markov với hữu hạn trạng thái

Biểu đồ chuyển trạng thái của một chuỗi Markov hữu hạn trạng thái là biểu đồ có:

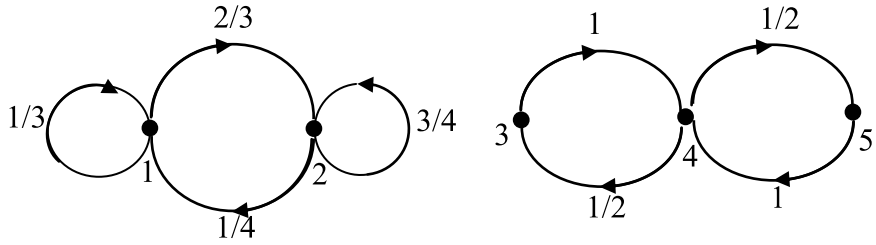
- Các đỉnh tương ứng với các trạng thái;

- Các đường nối giữa hai đỉnh i, j có mũi tên theo hướng từ i đến j nếu $p_{ij} > 0$.

Trong biểu đồ này nếu có thể di chuyển theo chiều mũi tên từ i đến j thì $i \to j$.

Biểu đồ chuyển trạng thái rất hữu ích trong việc xác định chuỗi Markov có không gian trạng thái tối giản hay không hoặc kiểm tra tính chất tuần hoàn và tìm chu kỳ của trạng thái.

Chuỗi Markov ở ví dụ 6.9 có biểu đồ:



Hình 6.4: Biểu đồ chuyển trạng thái

Các trạng thái có chu kỳ như sau:

$$d(1) = d(2) = 1;\ d(3) = d(4) = d(5) = 2.$$

Ví dụ này cho thấy các trạng thái cùng lớp sẽ có chu kỳ bằng nhau. Định lý sau đây khẳng định điều đó.

Định lý 6.7: Nếu $i \leftrightarrow j$ thì $d(i) = d(j)$. Do đó các trạng thái thuộc cùng một lớp có cùng chu kỳ.

Ví dụ 6.10: Vẽ biểu đồ chuyển trạng thái và phân loại các trạng thái của các chuỗi Markov có ma trận xác suất chuyển sau:

(a) $P=\begin{bmatrix} 0 & 0,5 & 0,5 \\ 0,5 & 0 & 0,5 \\ 0,5 & 0,5 & 0 \end{bmatrix}$

(b) $P=\begin{bmatrix} 0 & 0 & 0,5 & 0,5 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$

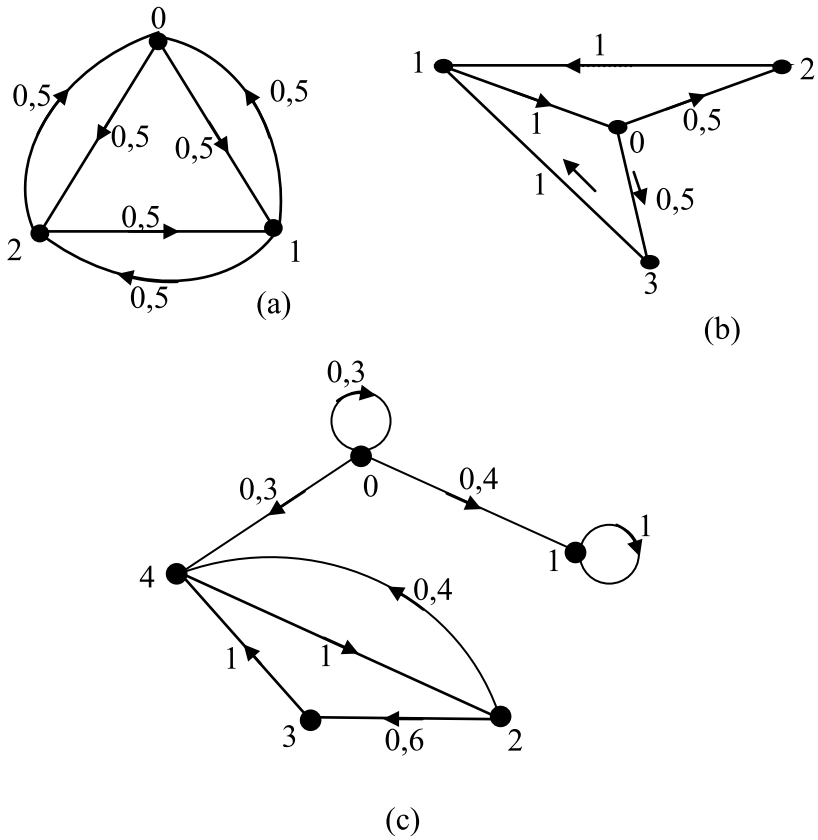
(c) $P=\begin{bmatrix} 0,3 & 0,4 & 0 & 0 & 0,3 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,6 & 0,4 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$

Giải: a) Biểu đồ chuyển trạng thái của chuỗi Markov (a) là hình 6.5 (a). Không gian trạng thái tối giản và không tuần hoàn. Chẳng hạn có thể xuất phát từ 0 và sau hai bước sẽ quay lại 0 theo đường $0\to 1\to 0$. Tuy nhiên cũng có thể xuất phát từ 0 và sau ba bước quay lại 0 theo đường $0\to 1\to 2\to 0$. Vậy 0 là trạng thái không tuần hoàn, do đó các trạng thái 1, 2 cũng không tuần hoàn.

b) Biểu đồ chuyển trạng thái của chuỗi Markov (b) là hình 6.5 (b). Không gian trạng thái tối giản và tuần hoàn với chu kỳ 3.

c) Biểu đồ chuyển trạng thái của chuỗi Markov (c) là hình 6.5 (c). Không gian trạng thái không tối giản vì trạng thái 0 và 4 không liên thông, $d(0)=d(1)=d(2)=d(4)=d(3)=1$.

Trạng thái i được gọi là *trạng thái hấp thụ* (Absorbing state) nếu $p_{ii}=1$, đó là trạng thái mà nếu hệ đạt đến trạng thái này thì không bao giờ rời đi. Trạng thái 1 của chuỗi Markov 6.3 (c) là trạng thái hấp thụ.



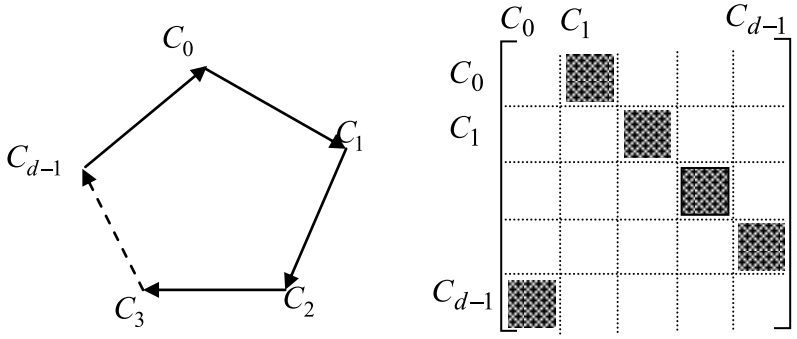
Hình 6.5: Biểu đồ chuyển trạng thái

6.3.4 Dạng ma trận xác suất chuyển của chuỗi Markov tối giản

Đối với chuỗi Markov tối giản mọi trạng thái đều có cùng chu kỳ, ta gọi d là chu kỳ chung của các trạng thái của chuỗi.

- Nếu $d = 1$ (không tuần hoàn) thì ma trận xác suất chuyển P chỉ có 1 khối.

- Nếu $d > 1$ thì tập trạng thái E tách thành d lớp con: $C_0, C_1, ..., C_{d-1}$. Trong trường hợp này sau một bước hệ xuất phát từ C_0 sẽ chuyển sang C_1; xuất phát từ C_1 sẽ chuyển sang C_2; v.v…; C_{d-1} sẽ chuyển sang C_0. Ma trận xác suất chuyển P có dạng khối như sau:



(6.37)

Vì vậy mỗi lớp con C_k có thể lấy làm không gian trạng thái của chuỗi Markov mới với ma trận xác suất chuyển là $\left[p_{ij}^{(d)}\right]_{i,j\in C_k}$, chuỗi này tối giản có chu kỳ bằng 1. Như vậy ta có thể quy việc nghiên cứu chuỗi Markov tổng quát (đặc biệt là vấn đề tìm $\lim\limits_{n\to\infty} p_{ij}^{(n)}$) về chuỗi Markov tối giản có chu kỳ 1.

6.3.5 Trạng thái hồi quy và trạng thái không hồi quy

Với mỗi trạng thái j ta gọi T_j là thời điểm (hoặc số các bước) lần đầu tiên hệ đến được trạng thái j sau thời điểm 0. Nếu hệ không bao giờ đến được trạng thái j ta đặt $T_j = \infty$. Vậy T_j là biến ngẫu nhiên rời rạc nhận giá trị trong tập $\{1,2,...,\infty\}$.

Với mỗi trạng thái i ta đặt:

$$\begin{aligned} f_{ij}^{(n)} &= P\left\{T_j = n \middle| X_0 = i\right\} \\ &= P\left\{X_n = j, X_1 \neq j, \ldots X_{n-1} \neq j \middle| X_0 = i\right\} \end{aligned} \tag{6.38a}$$

$$f_{ij}^{(0)} = 0 \text{ vì } T_j \geq 1. \tag{6.38b}$$

Như vậy $f_{ij}^{(n)}$ là xác suất để hệ xuất phát từ i lần đầu tiên chuyển sang j tại bước thứ n. Đặc biệt $f_{ii}^{(n)}$ là xác suất để hệ xuất phát từ i lần đầu tiên quay về i tại bước thứ n.

Các xác suất $f_{ij}^{(n)}$ có các tính chất sau:

$$f_{ij}^{(1)} = P\left\{T_j = 1 \middle| X_0 = i\right\} = P\left\{X_1 = j \middle| X_0 = i\right\} = p_{ij}. \qquad (6.39)$$

Từ tính Markov và công thức xác suất đầy đủ ta có:

$$f_{ij}^{(n)} = \sum_{k \neq j} p_{ik} f_{kj}^{(n-1)}, \ n \geq 2 \qquad (6.40)$$

$$p_{ij}^{(n)} = \sum_{k=0}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}, \quad n \geq 1 \qquad (6.41)$$

Xác suất để hệ xuất phát từ i chuyển đến trạng thái j tại thời điểm hữu hạn nào đó là:

$$f_{ij} = \sum_{n=0}^{\infty} f_{ij}^{(n)} = P\left\{T_j < \infty \middle| X_0 = i\right\} \qquad (6.42)$$

Xác suất để hệ xuất phát từ i quay trở về trạng thái i tại thời điểm hữu hạn nào đó là:

$$f_{ii} = \sum_{n=0}^{\infty} f_{ii}^{(n)} = P\left\{T_i < \infty \middle| X_0 = i\right\} \qquad (6.43)$$

Định nghĩa 6.9:

- Nếu $f_{ii} = P\left\{T_i < \infty \middle| X_0 = i\right\} = 1$ *thì* i *được gọi là trạng thái hồi quy (recurrent).*

- Nếu $f_{ii} = P\left\{T_i < \infty \middle| X_0 = i\right\} < 1$ *thì* i *được gọi là trạng thái không hồi quy (transient hoặc unrecurrent).*

Như vậy trạng thái i hồi quy khi và chỉ khi hệ xuất phát từ i, với xác suất 1 hệ lại trở về i tại thời điểm hữu hạn nào đó. Đối với trạng thái i không hồi quy thì "biến cố hệ không bao giờ quay trở lại i" có xác suất > 0 (xác suất $P\left\{T_i = \infty \middle| X_0 = i\right\} > 0$).

Trường hợp trạng thái i hồi quy ($f_{ii} = 1$), khi đó kỳ vọng của $(T_i \mid X_0 = i)$ (thời điểm trung bình để lần đầu tiên hệ quay lại trạng thái

i với điều kiện hệ xuất phát từ i tại thời điểm 0) được ký hiệu và tính theo công thức sau.

$$\mu_i = \mathrm{E}\left[T_i \middle| X_0 = i\right] = \sum_{n=0}^{\infty} n f_{ii}^{(n)} \tag{6.44}$$

Định nghĩa 6.10: *Giả sử i là trạng thái hồi qui, ta nói:*

- *i là trạng thái hồi quy dương (positive recurrent) nếu $\mu_i < \infty$.*
- *i là trạng thái hồi quy không (null recurrent) nếu $\mu_i = \infty$.*

Như vậy không gian trạng thái E được phân loại như sau:

1. Các trạng thái hồi quy:

- Trạng thái hồi quy dương;
- Trạng thái hồi quy không.

2. Các trạng thái không hồi quy.

Ví dụ 6.11: Xét chuỗi Markov với không gian trạng thái $E = \{0,1\}$ có ma trận xác suất chuyển:

$$P = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}.$$

Ta có: $f_{00}^{(1)} = p_{00} = 1 \qquad f_{10}^{(1)} = p_{10} = \dfrac{1}{2}$

$$f_{00}^{(2)} = p_{01} f_{10}^{(1)} = 0\left(\frac{1}{2}\right) = 0$$

$$f_{00}^{(n)} = 0, \quad \forall n \geq 2$$

Theo công thức (6.42)

$$f_{00} = P\left\{T_0 < \infty \middle| X_0 = 0\right\} = \sum_{n=0}^{\infty} f_{00}^{(n)} = 1 + 0 + 0 + \ldots = 1$$

$$\mu_0 = \mathrm{E}\left[T_0 \middle| X_0 = 0\right] = \sum_{n=0}^{\infty} n f_{00}^{(n)} = 1$$

Như vậy 0 là trạng thái hồi quy dương.

Tương tự ta có:

$$f_{11}^{(1)} = p_{11} = \frac{1}{2}$$

$$f_{01}^{(1)} = p_{01} = 0$$

$$f_{11}^{(2)} = p_{10} f_{01}^{(1)} = \left(\frac{1}{2}\right)0 = 0$$

$$f_{11}^{(n)} = 0, \quad \forall n \geq 2$$

Theo công thức (6.42)

$$f_{11} = P\left\{T_1 < \infty \middle| X_0 = 1\right\} = \sum_{n=1}^{\infty} f_{00}^{(n)} = \frac{1}{2} + 0 + 0 + \ldots = \frac{1}{2} < 1$$

Như vậy 1 là trạng thái không hồi quy.

6.3.6 Tiêu chuẩn hồi quy và không hồi quy

Định lý 6.8:

1) Trạng thái i là hồi quy khi và chỉ khi $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

2) Trạng thái i là không hồi quy khi và chỉ khi $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.

3) Nếu $i \to j$ và i hồi quy thì $j \to i$ và j cũng hồi quy.

4) Nếu $i \leftrightarrow j$ và j hồi quy thì $f_{ij} = 1$.

Định lý 6.9: Nếu j không hồi quy thì với mọi trạng thái i ta có:

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty.$$

Đặc biệt: $\lim_{n\to\infty} p_{ij}^{(n)} = 0$ với mọi $i \in E$ (6.45)

6.3.7 Định lý giới hạn cơ bản của chuỗi Markov

Định lý 6.10: Giả sử j là trạng thái hồi quy không tuần hoàn ($d(j)=1$). Khi đó:

1. Nếu i và j liên thông thì:

$$\lim_{n\to\infty} p_{ij}^{(n)} = \begin{cases} \dfrac{1}{\mu_j} & \text{đối với } j \text{ là trạng thái dương} \\ 0 & \text{đối với } j \text{ là trạng thái không} \end{cases} \tag{6.46}$$

2. Nếu i và j không liên thông thì:

$$\lim_{n\to\infty} p_{ij}^{(n)} = \begin{cases} \dfrac{f_{ij}}{\mu_j} & \text{đối với } j \text{ là trạng thái dương} \\ 0 & \text{đối với } j \text{ là trạng thái không} \end{cases} \tag{6.47}$$

Định lý 6.11: Giả sử j là trạng thái hồi quy tuần hoàn chu kỳ $d(j)=d>1$. Khi đó:

1. Nếu i và j liên thông; i thuộc vào lớp con C_r còn j thuộc vào lớp C_{r+a} thì:

$$\lim_{n\to\infty} p_{ij}^{(nd+a)} = \frac{d}{\mu_j},\ (a=0,1,...,d-1) \tag{6.48}$$

2. Nếu i bất kỳ thì:

$$\lim_{n\to\infty} p_{ij}^{(nd+a)} = \left[\sum_{r=0}^{\infty} f_{ij}^{(rd+a)}\right]\frac{d}{\mu_j},\ (a=0,1,...,d-1) \tag{6.49}$$

6.3.8 Sự tồn tại phân bố dừng và phân bố giới hạn

Định lý 6.12: Giả sử $\{X_n\}$ là chuỗi Markov với không gian trạng thái E. Khi đó phân bố dừng tồn tại và duy nhất khi và chỉ khi trong số các lớp tương đương của không gian trạng thái E có đúng một lớp hồi quy dương.

Giả sử lớp hồi quy dương là C, khi đó phân bố dừng có dạng

$$p_j = \begin{cases} \dfrac{1}{\mu_j} & \text{nếu} \quad j \in C \\ 0 & \text{nếu} \quad j \notin C. \end{cases} \tag{6.50}$$

Định lý 6.13: Điều kiện cần và đủ để tồn tại phân bố giới hạn là không gian trạng thái E có đúng một lớp hồi quy dương C không tuần hoàn ($d(\mathrm{C}) = 1$) sao cho $f_{ij} = 1;\ \forall j \in \mathrm{C},\ \forall i \in E$.

Khi đó phân bố giới hạn cũng là phân bố dừng duy nhất thỏa mãn công thức (6.50).

Định lý 6.14: Giả sử $\{X_n, n \geq 0\}$ là một chuỗi Markov có không gian trạng thái hữu hạn. Khi đó các điều sau là tương đương:

(i) $\{X_n, n \geq 0\}$ tối giản, không tuần hoàn.

(ii) $\{X_n, n \geq 0\}$ tối giản, không tuần hoàn và tất cả các trạng thái là hồi quy dương.

(iii) $\{X_n, n \geq 0\}$ có tính ergodic, nghĩa là tồn tại phân bố ergodic (thỏa mãn công thức 6.32, 6.34).

(iv) Tồn tại n_0 sao cho $\min\limits_{i,j} p_{ij}^{(n)} > 0$ với mọi $n \geq n_0$ (xem định lý 6.4).

6.4. DI ĐỘNG NGẪU NHIÊN TRÊN ĐƯỜNG THẲNG

6.4.1 Di động ngẫu nhiên trên đường thẳng không có trạng thái hấp thu

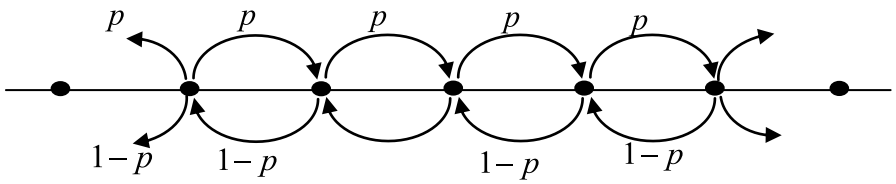
Giả sử $\{\varepsilon_n\}_{n=1}^{\infty}$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố thỏa mãn:

$$P\{\varepsilon_n = 1\} = p,\ P\{\varepsilon_n = -1\} = 1 - p = q;\ 0 < p < 1. \tag{6.51}$$

Đặt $X_n = \varepsilon_1 + \varepsilon_2 + \cdots \varepsilon_n$, $(n = 1, 2, ...)$. Khi đó $\{X_n\}_{n=1}^{\infty}$ lập thành chuỗi Markov với ma trận xác suất chuyển là:

$$P=\left[p_{ij}\right] \text{ trong đó } p_{ij}=\begin{cases} p & \text{với } j=i+1 \\ 1-p & \text{với } j=i-1 \\ 0 & \text{với } j\neq i\pm 1 \end{cases} \tag{6.52}$$

Không gian trạng thái của chuỗi này là $E=\{0,\pm 1,\pm 2,...\}$. Biểu đồ chuyển trạng thái có dạng sau:



Hình 6.6: Di động ngẫu nhiên trên đường thẳng không có trạng thái hấp thu

Chuỗi này dùng để mô tả *di động ngẫu nhiên trên đường thẳng* của hạt vật chất nào đó (có tài liệu gọi là du động ngẫu nhiên): Sau mỗi chu kỳ hạt dịch chuyển sang phải với xác suất p hoặc dịch sang trái với xác suất $1-p$.

Di động ngẫu nhiên trên đường thẳng là chuỗi Markov tối giản, có chu kỳ $d=2$.

- Nếu $p=q=\dfrac{1}{2}$ thì mọi trạng thái là hồi quy không.

- Nếu $p\neq q$ thì mọi trạng thái là không hồi quy.

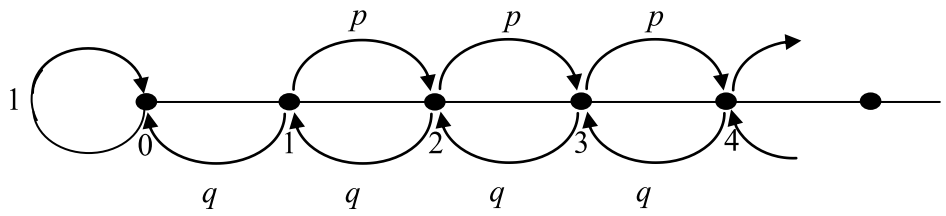
Theo định lý 6.12 chuỗi không tồn tại phân bố dừng và do đó không có tính ergodic.

6.4.2 Di động ngẫu nhiên trên đường thẳng có một trạng thái hấp thu

Đó là di động của hạt vật chất với không gian trạng thái $E=\{0,1,2,...\}$ và ma trận xác suất chuyển là: $P=\left[p_{ij}\right]$,

$$\text{trong đó } p_{00}=1,\ p_{ij}=\begin{cases} p & \text{với } j=i+1, i\neq 0 \\ 1-p & \text{với } j=i-1, i\neq 0 \\ 0 & \text{với } j\neq i\pm 1, i\neq 0 \end{cases};\ 0<p<1 \tag{6.53}$$

Biểu đồ chuyển trạng thái có dạng:



Hình 6.7: Di động ngẫu nhiên trên đường thẳng có một trạng thái hấp thu

Lúc này $\{0\}$ lập thành lớp hồi quy dương duy nhất với chu kỳ $d=1$: $E=\{0\}\cup\{1,2,...\}$. Tất cả các trạng thái $1,2,...$ là không hồi quy (vì nếu 1 hồi quy chẳng hạn thì do $1\to 0$ nên $0\to 1$, điều này không thể xảy ra). Vì vậy theo định lý 6.12 và công thức (6.50) tồn tại phân bố dừng duy nhất, đó là:

$$p_j=\begin{cases}1 & \text{với } j=0,\\ 0 & \text{với } j\neq 0.\end{cases} \tag{6.54}$$

Hơn nữa có thể chứng minh được rằng đối với mỗi $i\geq 1$ thì:

$$\lim_{n\to\infty} p_{i0}^{(n)}=\begin{cases}(q/p)^i & \text{nếu } p>q,\\ 1 & \text{nếu } p\leq q;\ q=1-p.\end{cases} \tag{6.55}$$

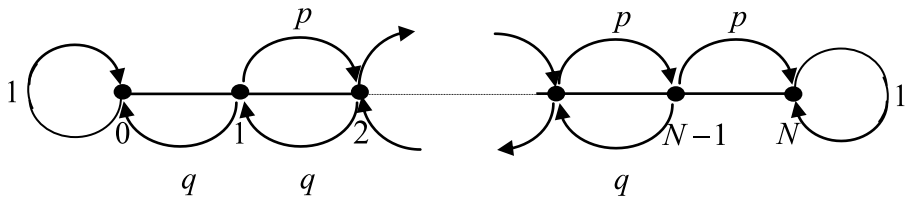
Vì vậy:

- Khi $p>q$ thì $\lim\limits_{n\to\infty} p_{i0}^{(n)}=(q/p)^i$ phụ thuộc vào i, do đó không tồn tại phân bố giới hạn.

- Khi $p\leq q$ thì tồn tại phân bố giới hạn, đó là $p_j=\begin{cases}1 & \text{với } j=0,\\ 0 & \text{với } j\neq 0.\end{cases}$

6.4.3 Di động ngẫu nhiên trên đường thẳng có hai trạng thái hấp thu

Di động ngẫu nhiên trên đường thẳng với hai trạng thái hấp thụ có biểu đồ chuyển trạng thái:



Hình 6.8: Di động ngẫu nhiên trên đường thẳng có hai trạng thái hấp thu

Ma trận xác suất chuyển là ma trận vuông cấp $N+1$ có dạng

$$P=\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & \cdots & \cdots & 0 \\ q & 0 & p & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & q & 0 & p & \cdots & \cdots & \cdots & 0 \\ \vdots & & & & \ddots & & & \vdots \\ \vdots & & & & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}$$

Trong trường hợp này có hai lớp hồi quy dương $\{0\}$ và $\{N\}$. Các trạng thái còn lại không hồi quy:

$$E=\{0\}\cup\{N\}\cup\{1,2,...,N-1\}.$$

0, N là hai trạng thái hấp thụ do đó hệ sẽ ngừng lại khi hạt rơi vào trạng thái 0 hoặc trạng thái N.

Vì vậy tồn tại vô số phân bố dừng $\mathbf{P}^*=[p_0,p_1,...,p_N]$, trong đó:

$$p_0=a,\ p_N=1-a,\ p_1=p_2=...=p_{N-1}=0,\ \text{với}\ 0\le a\le 1. \quad (6.56)$$

Có thể chứng minh được rằng:

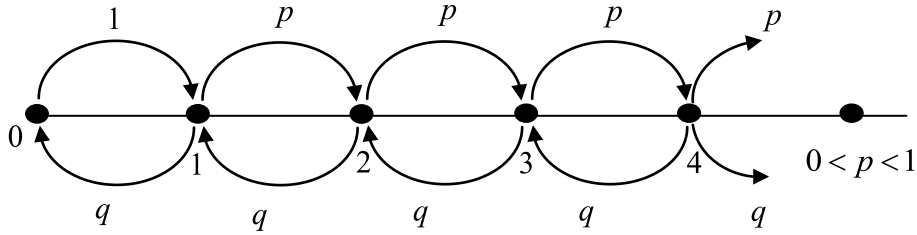
$$\lim_{n\to\infty}p_{i0}^{(n)}=\begin{cases}\dfrac{(q/p)^i-(q/p)^N}{1-(q/p)^N} & \text{nếu } p\ne q,\\ 1-\dfrac{i}{N} & \text{nếu } p=q=1/2.\end{cases}$$

$$\lim_{n\to\infty} p_{iN}^{(n)} = 1 - \lim_{n\to\infty} p_{i0}^{(n)},\ \lim_{n\to\infty} p_{ij}^{(n)} = 0\ \ (\forall j = 1,2,...,N-1).$$

Vì vậy không tồn tại phân bố giới hạn.

6.4.4 Di động ngẫu nhiên trên đường thẳng có một trạng thái phản hồi

Di động ngẫu nhiên trên đường thẳng với một trạng thái phản hồi có biểu đồ chuyển trạng thái



Hình 6.9: Di động ngẫu nhiên trên đường thẳng có một trạng thái phản hồi

Không gian trạng thái $E = \{0,1,2,...\}$. Chuỗi tối giản, có chu kỳ $d = 2$ do đó không tồn tại phân bố giới hạn.

Ma trận xác suất chuyển

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & \cdots \\ q & 0 & p & 0 & 0 & \cdots & \cdots \\ 0 & q & 0 & p & 0 & & \\ 0 & 0 & q & 0 & \ddots & & \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \\ \vdots & \vdots & & & \ddots & \ddots & \\ \vdots & \vdots & & & & & \end{bmatrix}$$

Hệ phương trình xác định phân bố dừng (6.35) có dạng:

$$\begin{cases} x_0 = x_1 q \\ x_1 = x_0 + x_2 q \\ x_2 = x_1 p + x_3 q \\ \ldots\ldots\ldots\ldots \\ x_i \geq 0,\ \sum x_i = 1 \end{cases} \text{ hay là } \begin{cases} x_0 = x_1 q \\ x_1 = x_1 q + x_2 q \\ x_2 = x_2 q + x_3 q \\ \ldots\ldots\ldots\ldots \\ x_i \geq 0,\ \sum x_i = 1 \end{cases} \quad (6.57)$$

Từ đó suy ra:

$$x_1 = \frac{x_0}{q},\ x_j = (p/q)x_{j-1} = \ldots = (p/q)^{j-1} x_1,\ \forall j = 2,3,\ldots \quad (6.58)$$

- Khi $p > q$, hạt có xu hướng đi sang phải, tất cả các trạng thái không hồi quy. Nghiệm (6.58) không thỏa mãn điều kiện $\sum x_i = 1$. Chuỗi không tồn tại phân bố dừng.

- Khi $p = q = 1/2$, tất cả các trạng thái là hồi quy không. Nghiệm (6.58) không thỏa mãn điều kiện $\sum x_i = 1$. Không tồn tại phân bố dừng.

- Khi $p < q$, tất cả các trạng thái là hồi quy dương.

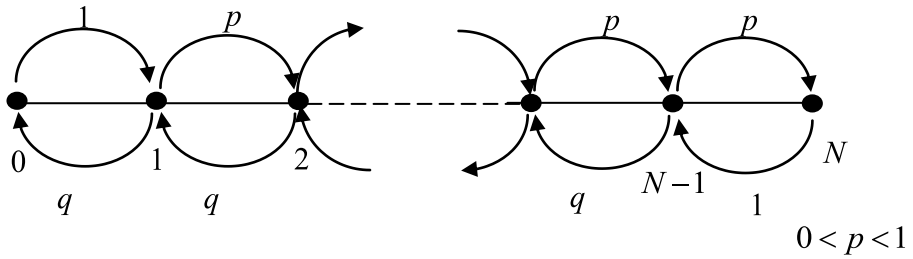
$$1 = \sum_{i=0}^{\infty} x_i = x_0 + \frac{x_0}{q} \sum_{i=0}^{\infty} (p/q)^i = x_0 + \frac{x_0}{q-p} = \frac{2qx_0}{q-p}$$

Do đó chuỗi tồn tại phân bố dừng duy nhất.

$$p_0 = \frac{q-p}{2q}, p_1 = \frac{q-p}{2q^2}, \ldots, p_j = \frac{q-p}{2q^2}\left(\frac{p}{q}\right)^{j-1};\ j \geq 2 \quad (6.59)$$

6.4.5 Di động ngẫu nhiên trên đường thẳng có hai trạng thái phản hồi

Di động ngẫu nhiên trên đường thẳng với hai trạng thái phản hồi có biểu đồ chuyển trạng thái:



Hình 6.10: Di động ngẫu nhiên trên đường thẳng có hai trạng thái phản hồi

Chuỗi tối giản, không gian trạng thái hữu hạn $E=\{0,1,2,...,N\}$. Tất cả các trạng thái của chuỗi là hồi quy dương, có chu kỳ $d=2$ do đó không tồn tại phân bố giới hạn.

Ma trận xác suất chuyển:

$$P=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & 0 & \cdots & 0 \\ 0 & q & 0 & p & 0 & & 0 \\ 0 & 0 & q & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}$$

Phân bố dừng là nghiệm duy nhất của hệ phương trình (6.35)

$$\begin{cases} x_0 = x_1 q \\ x_1 = x_0 + x_2 q \\ x_2 = x_1 p + x_3 q \\ ..... \\ x_{N-1} = x_{N-2} p + x_N \\ x_N = x_{N-1} p \\ x_i \geq 0,\ \sum_{i=0}^{N} x_i = 1 \end{cases} \qquad (6.60)$$

Từ đó suy ra phân bố dừng:

$$p_0 = \frac{1}{1+\left(p/q\right)^{N-1} + \dfrac{1}{q}\displaystyle\sum_{j=1}^{N-1}\left(p/q\right)^{j-1}};$$

$$p_i = \frac{p_0}{q}\left(\frac{p}{q}\right)^{i-1}, \; 1 \le i \le N-1; \tag{6.61}$$

$$p_N = p_0\left(\frac{p}{q}\right)^{N-1}$$

Chuỗi tồn tại phân bố dừng nhưng không tồn tại phân bố giới hạn.

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

6.1 Quá trình ngẫu nhiên $\{X(t); t \in I\}$ là một hàm số của biến số t.

Đúng ☐ Sai ☐ .

6.2 Mọi quá trình có gia số độc lập là quá trình Markov.

Đúng ☐ Sai ☐ .

6.3 Chuỗi Markov là quá trình Markov $\{X(t); t \in I\}$ có không gian trạng thái E đếm được.

Đúng ☐ Sai ☐ .

6.4 Ma trận xác suất chuyển sau n bước của một chuỗi Markov bằng tích *n* lần ma trận xác suất chuyển một bước của chuỗi Markov này.

Đúng ☐ Sai ☐ .

6.5 Nếu tồn tại phân bố giới hạn thì nó là phân bố dừng duy nhất.

Đúng ☐ Sai ☐ .

6.6 Mọi chuỗi Markov có hữu hạn trạng thái luôn tồn tại phân bố dừng duy nhất đó là phân bố ergodic.

Đúng ☐ Sai ☐ .

6.7 Cho chuỗi Markov $\{X_n\}_{n=1}^{\infty}$ với không gian trạng thái $E = \{0,1,2\}$ và ma trận xác suất chuyển

$$P = \begin{bmatrix} 0,1 & 0,2 & 0,7 \\ 0,9 & 0,1 & 0,0 \\ 0,1 & 0,8 & 0,1 \end{bmatrix}$$

Biết phân bố ban đầu:

$$p_0 = P\{X_0 = 0\} = 0,3;$$

$$p_1 = P\{X_0 = 1\} = 0,4;$$

$p_2 = P\{X_0 = 2\} = 0,3$.

Tính $P\{X_0 = 0, X_1 = 2, X_2 = 1\}$.

6.8 Giả sử $P = \left[p_{ij}\right] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$ là một ma trận Markov

(là ma trận thỏa mãn điều kiện $p_{ij} \geq 0$; $\sum_{j=1}^{m} p_{ij} = 1$, $\forall i = 1,...,m$).

Chứng minh rằng P^n cũng là ma trận Markov, với mọi số tự nhiên dương n.

6.9 Cho chuỗi Markov $\{X_n\}_{n=1}^{\infty}$ với không gian trạng thái $E = \{0,1,2\}$ và ma trận xác suất chuyển

$$P = \begin{bmatrix} 0,1 & 0,2 & 0,7 \\ 0,2 & 0,2 & 0,6 \\ 0,6 & 0,1 & 0,3 \end{bmatrix}$$

a) Tính ma trận xác suất chuyển 2 bước.

b) Tính $P\{X_3 = 1 | X_1 = 0\}$; $P\{X_3 = 1 | X_0 = 0\}$.

c) Tìm phân bố dừng.

6.10 Xét bài toán truyền một bức điện gồm các tín hiệu 0, 1 thông qua kênh có nhiều trạm và mỗi trạm nhận sai tín hiệu với xác suất không đổi bằng $\alpha \in (0,1)$. Giả sử X_0 là tín hiệu truyền đi và X_n là tín hiệu nhận được tại trạm n. Cho biết $\{X_n; n = 0,1,2,...\}$ lập thành chuỗi Markov với ma trận xác suất chuyển:

$$P = \begin{bmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{bmatrix}.$$

Giả sử hai tín hiệu 0, 1 được truyền đi đồng khả năng.

a) Tính $P\{X_0 = 0, X_1 = 0, X_2 = 0\}$.

b) Tính $P\{X_0 = 0, X_1 = 0, X_2 = 0\} + P\{X_0 = 0, X_1 = 1, X_2 = 0\}$.

c) Tính $P\{X_5 = 0 | X_0 = 0\}$.

6.11 Xét chuỗi Markov với không gian trạng thái $E = \{a, b, c, d\}$ và ma trận xác suất chuyển:

$$P = \begin{bmatrix} 0,1 & 0,3 & 0,2 & 0,4 \\ 0,2 & 0.3 & 0,2 & 0,3 \\ 0,3 & 0,5 & 0,1 & 0,1 \\ 0,2 & 0,1 & 0,4 & 0,3 \end{bmatrix}$$

a) Tìm xác suất chuỗi đi theo đường đi: $b - a - b - c - b - a$.

b) Tính xác suất $P\{X_1 = a, X_3 = c, X_4 = b, X_5 = a | X_0 = b\}$.

c) Tính $P^{(5)}$.

d) Tính $P\{X_5 = a | X_0 = b\}$.

e) Tìm $\mathbf{P}(5)$ biết $\mathbf{P}(0) = [0,2 \quad 0,4 \quad 0,3 \quad 0,1]$.

6.12 Xét chuỗi Markov với không gian trạng thái $E = \{0,1\}$ và ma trận xác suất chuyển:

$$P = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}, \ 0 < a < 1, \ 0 < b < 1.$$

Giả sử phân bố đầu $P\{X_0 = 0\} = P\{X_0 = 1\} = 0,5$.

a) Vẽ biểu đồ chuyển trạng thái.

b) Tìm phân bố dừng.

c) Tìm phân bố của X_n.

d) Tìm phân bố giới hạn khi $a + b \le 1$.

6.13 Xét mô hình kiểm kê phụ tùng thay thế với $s=0$ và $S=3$ là các mức căn cứ để nhập hàng cùng với ξ_n là lượng hàng khách yêu cầu trong chu kỳ n. Biết rằng:

$$P\{\xi_n=0\}=0{,}4;\ P\{\xi_n=1\}=0{,}3;\ P\{\xi_n=2\}=0{,}3 \text{ với mọi } n.$$

Xác định ma trận xác suất chuyển của chuỗi Markov $\{X_n\}$, trong đó X_n là số phụ tùng còn lại tại cuối chu kỳ n.

6.14 Hai công ty *A* và *B* cung cấp cho thị trường cùng một loại sản phẩm. Hiện tại công ty *A* chiếm 60% và công ty *B* chiếm 40% thị phần. Mỗi năm *A* mất 2/3 thị phần của mình cho *B* và *B* mất 1/2 thị phần cho *A*. Tìm tỉ lệ thị phần hai công ty chiếm được sau hai năm.

6.15 Mỗi một người dân của thị trấn *N* có một trong ba nghề (*A, B, C*). Con cái họ nối tiếp nghề của cha mình với xác suất tương ứng là $(3/5,\ 2/3,\ 1/4)$. Nếu không theo nghề của cha thì chúng chọn một trong hai nghề còn lại với xác suất như nhau. Hãy tìm:

a) Phân bố theo nghề nghiệp của dân cư thị trấn ở thế hệ tiếp theo, nếu thế hệ hiện tại có tỉ lệ theo nghề nghiệp là 20% có nghề *A*, 30% có nghề *B* và 50% có nghề *C*.

b) Phân bố giới hạn theo nghề nghiệp của dân cư thị trấn ở thế hệ tương lai xa.

6.16 Chứng minh rằng $\leftrightarrow$ là một quan hệ tương đương trong tập các trạng thái.

6.17 Cho chuỗi Markov có không gian trạng thái $E=\{0,1,2\}$ và ma trận xác suất chuyển:

$$P=\begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

a) Vẽ biểu đồ chuyển trạng thái.

b) Tính $P^{(2n)}$, $P^{(2n+1)}$.

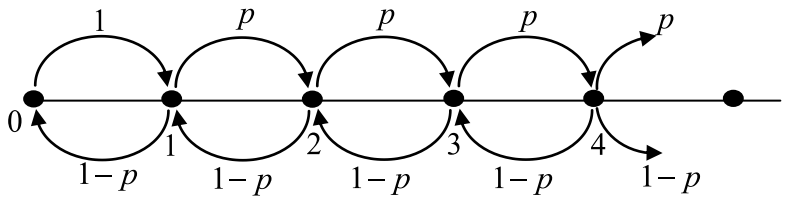
c) Chứng tỏ rằng 0 là trạng thái tuần hoàn có chu kỳ 2.

6.17 Cho chuỗi Markov có không gian trạng thái $E=\{0,1,2,3,4\}$ và ma trận xác suất chuyển:

$$P=\begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 0 & 1/4 & 1/4 \end{bmatrix}$$

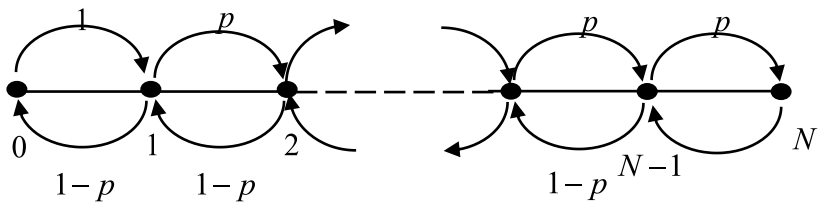
Tìm các lớp liên thông của không gian trạng thái. Vẽ biểu đồ trạng thái và tìm chu kỳ các trạng thái.

6.18 Xét di động ngẫu nhiên trên đường thẳng có một trạng thái phản hồi có dạng như hình vẽ



Với $p=0,4$. Tìm phân bố dừng.

6.19 Xét di động ngẫu nhiên trên đường thẳng có hai trạng thái phản hồi có dạng như hình vẽ:



Với $N=10;\ p=0,6$. Tìm phân bố dừng.

HƯỚNG DẪN GIẢI BÀI TẬP

Đáp án chương 1

Câu hỏi	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10
Đáp án	Đúng	Sai	Đúng	Đúng	Sai	Đúng	Sai	Sai	Đúng	Đúng

1.11 a) $P = 0{,}246$ b) $P = 0{,}495$.

1.12 Mỗi khách đều có 6 khả năng để ra ở 6 tầng còn lại của tòa nhà. Do đó số kết cục đồng khả năng có thể $N = 6^3 = 216$. Gọi A là biến cố tất cả cùng ra ở tầng bốn, biến cố này chỉ có 1 trường hợp thuận lợi. Do đó $P(A) = \dfrac{1}{216}$.

Lý luận tương tự trên ta có $P(B) = \dfrac{6}{216} = \dfrac{1}{36}$; $P(C) = \dfrac{A_6^3}{216} = \dfrac{5}{9}$.

1.13 $P = \dfrac{1}{720}$.

1.15 Gọi A_1 và A_2 tương ứng là biến cố người thứ nhất và thứ hai bắn trúng mục tiêu, A là biến cố chỉ có một người bắn trúng mục tiêu. $A = A_1\overline{A}_2 + \overline{A}_1A_2$. Sử dụng qui tắc cộng xác suất trường hợp xung khắc và qui tắc nhân trường hợp độc lập ta có:

$$P(A) = P(A_1\overline{A}_2) + P(\overline{A}_1A_2) = P(A_1)P(\overline{A}_2) + P(\overline{A}_1)P(A_2)$$
$$= 0{,}8 \cdot 0{,}1 + 0{,}2 \cdot 0{,}9 = 0{,}26 .$$

Tương tự ta có: $P_b = 0{,}98$; $P_c = 0{,}02$.

1.16 Gọi A_1 là biến cố sản phẩm lấy ra thuộc loại 1.

Gọi A_2 là biến cố sản phẩm lấy ra thuộc loại 2.

Gọi A là biến cố sản phẩm lấy ra thuộc loại 1 hoặc loại 2: $A = A_1 + A_2$

Vì A_1, A_2 xung khắc do đó:

$$P(A) = P(A_1 + A_2) = P(A_1) + P(A_2) = 0,4 + 0,5 = 0,9$$

1.17 $P = C_{30}^5 \left(\frac{1}{50}\right)^5 \left(\frac{49}{50}\right)^{25} = 0,00027$.

1.18 Gọi A_i là biến cố sản phẩm đã qua kiểm tra chất lượng ở phòng thứ i, i=1,2,3.

Gọi B là biến cố phế phẩm được nhập kho.

$$P(B) = P(\overline{A_1})P(\overline{A_2})P(\overline{A_3}) = (1-0,8)(1-0,9)(1-0,99) = 0,0002.$$

1.19 a) $\Omega = \{e_1, e_2, ...\}$, e_k là biến cố sơ cấp chỉ lần đầu tiên xuất hiện mặt sấp ở lần tung thứ k.

b) $P(e_k) = q^{k-1}p$; $q = 1-p$.

c) $P(\Omega) = \sum_{k=1}^{\infty} q^{k-1}p = \frac{p}{1-q} = 1$.

d)
$$P(B) = \sum_{m=1}^{\infty} P(e_{2m}) = \sum_{m=1}^{\infty} q^{2m-1}p = \frac{p}{q}\sum_{m=1}^{\infty} q^{2m}$$
$$= \frac{pq}{1-q^2} = \frac{q}{1+q} = \frac{1-p}{2-p}$$

1.20 $P = 0,11$.

1.21 Gọi A_i là biến cố lần thứ i lấy ra 3 sản phẩm mới để kiểm tra, ($i = \overline{1,3}$). Gọi A là biến cố sau 3 lần kiểm tra tất cả các sản phẩm đều được kiểm tra $A = A_1 A_2 A_3$. Vì các biến cố phụ thuộc nên

$$P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) = 1 \cdot \frac{5}{21} \cdot \frac{1}{84} = \frac{5}{1764}.$$

1.22 $P(A) + P(B) - P(A)P(B) = 0,65 \Rightarrow P(B) = 0,5$.

1.23 a) $P(A \mid B) = 3/4$

1.24 Gọi A_i là biến cố lần thứ nhất chọn được là số i, $i = 0,1,...,9$. Gọi B là biến cố lần thứ hai chọn được số 4.

$$P(A_i) = \frac{1}{10};\ P(B \mid A_i) = \frac{1}{9},\ \forall i \neq 4;\ P(B \mid A_4) = 0.$$

Vậy $P(B) = \sum_{i=1}^{10} P(A_i)P(B \mid A_i) = 9\left(\frac{1}{9}\right)\left(\frac{1}{10}\right) = \frac{1}{10}$.

1.25 Gọi A là biến cố sản phẩm kiểm tra là phế phẩm.

Gọi B_i là biến cố sản phẩm lấy ra kiểm tra thuộc phân xưởng thứ i, i = 1,2 3.

$P(B_1) = 0,36;\ P(B_2) = 0,34;\ P(B_3) = 0,30$. Hệ $\{B_1, B_2, B_3\}$ đầy đủ $P(A|B_1) = 0,12;\ P(A|B_2) = 0,10;\ P(A|B_3) = 0,08$.

a) $$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)$$
$$= 0,1012$$

b) $$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(A)} = \frac{0,36 \times 0,12}{0,1012} = 0,427$$

$$P(B_2|A) = \frac{P(B_2)P(A|B_2)}{P(A)} = \frac{0,34 \times 0,10}{0,1012} = 0,336$$

$$P(B_3|A) = \frac{P(B_3)P(A|B_3)}{P(A)} = \frac{0,30 \times 0,08}{0,1012} = 0,237$$

1.26 Gọi B_i là biến cố xạ thủ được xét thuộc nhóm thứ i, i = 1,2,3,4.

Gọi A là biến cố xạ thủ bắn trượt.

Theo đề bài ta có:

$$P(B_1)=\frac{5}{18},\ P(B_2)=\frac{7}{18},\ P(B_3)=\frac{4}{18},\ P(B_4)=\frac{2}{18}$$

$$P(A|B_1)=0,2,\ P(A|B_2)=0,3,\ P(A|B_3)=0,4,\ P(A|B_4)=0,5$$

$$P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)$$
$$+\ P(B_3)P(A|B_3)+\ P(B_4)P(A|B_4)$$
$$=\frac{5}{18}\times 0,2+\frac{7}{18}\times 0,3+\frac{4}{18}\times\ 0,4+\frac{2}{18}\times 0,5=\frac{57}{180}$$

Áp dụng công thức Bayes, ta thu được:

$$P(B_1|A)=\frac{P(B_1)P(A|B_1)}{P(A)}=\frac{\frac{5}{18}\times 0,2}{\frac{57}{180}}=\frac{10}{57},$$

tương tự: $P(B_2|A)=\frac{21}{57},\ P(B_3|A)=\frac{16}{57},\ P(B_4|A)=\frac{10}{57}$.

Vậy xạ thủ có khả năng ở nhóm thứ hai nhất.

1.27 Gọi B_1 là biến cố viên đạn thứ nhất trúng mục tiêu, $P(B_1)=0,7$.

Gọi B_2 là biến cố viên đạn thứ hai trúng mục tiêu, $P(B_2)=0,4$.

Hai biến cố này độc lập.

Xác suất biến cố chỉ có viên đạn thứ nhất trúng mục tiêu

a) $$P(A)=P\left(B_1\overline{B_2}\cup B_2\overline{B_1}\right)=P\left(B_1\overline{B_2}\right)+P\left(B_2\overline{B_1}\right)$$
$$=0,7\times 0,6+0,4\times 0,3=0,54$$

b) $P(B_1|A)=\dfrac{P(B_1A)}{P(A)}=\dfrac{P\left(B_1\left[B_1\overline{B_2}\cup B_2\overline{B_1}\right]\right)}{P(A)}$

$$=\frac{P\left(B_1\overline{B_2}\right)}{P(A)}=\frac{0,7\times 0,6}{0,54}=0,778$$

1.28 Gọi A là biến cố sản phẩm kiểm tra có kết luận đạt tiêu chuẩn chất lượng.

Gọi B_T là biến cố sản phẩm đạt tiêu chuẩn chất lượng.

Gọi B_H là biến cố sản phẩm không đạt tiêu chuẩn chất lượng.

$$P(B_T)=0,85;\ P(B_H)=0,15$$

Hệ $\{B_T,B_H\}$ đầy đủ:

$$P(A|B_T)=0,9;\ P(\overline{A}|B_H)=0,95$$
$$\Rightarrow\ P(\overline{A}|B_T)=0,1;\ P(A|B_H)=0,05.$$

a) $$P(A)=P(B_T)P(A|B_T)+P(B_H)P(A|B_H)$$
$$=0,85\times 0,9+0,15\times 0,05=0,7725$$

b) $P(B_H|A)=\dfrac{P(B_H)P(A|B_H)}{P(A)}=\dfrac{0,15\times 0,05}{0,7725}=0,0097$

c) $$P\left(AB_T\cup\overline{A}B_H\right)=P(AB_T)+P(\overline{A}B_H)$$
$$=0,85\times 0,9+0,15\times 0,95=0,9075.$$

1.29 Giả sử biến cố A xuất hiện lần đầu tiên ở lần thử thứ n. Nếu biến cố A xuất hiện trước biến cố B thì biến cố C phải xuất hiện ở $n-1$ phép thử đầu tiên. Đặt D là biến cố A xuất hiện trước B thì:

$$D=\bigcup_{n=1}^{\infty}D_n$$

Trong đó D_n là biến cố C phải xuất hiện ở $n-1$ phép thử đầu tiên và A xuất hiện ở lần thử thứ n. Vì các biến cố D_n xung khắc từng đôi một, do đó

$$P(D)=\sum_{n=1}^{\infty}P(D_n)=\sum_{n=1}^{\infty}\left[P(C)\right]^{n-1}P(A)$$
$$=\sum_{n=1}^{\infty}r^{n-1}p=p\sum_{n=0}^{\infty}r^n=\frac{p}{1-r}=\frac{p}{p+q}.$$

1.30 Nếu $P(A\,|\,B)=\dfrac{P(AB)}{P(B)}>P(A)$ thì $P(AB)>P(A)P(B)$.

Do đó $P(B\,|\,A)=\dfrac{P(AB)}{P(A)}>\dfrac{P(A)P(B)}{P(A)}=P(B)$.

Đáp án chương 2

Câu hỏi	**2.1**	**2.2**	**2.3**	**2.4**	**2.5**	**2.6**	**2.7**	**2.8**
Đáp án	Sai	Sai	Đúng	Sai	Đúng	Sai	Sai	Đúng
Câu hỏi	**2.9**	**2.10**	**2.11**	**2.12**	**2.13**	**2.14**	**2.15**	
Đáp án	Đúng	Sai	Sai	Sai	Đúng	Đúng	Sai	

2.16 $0\le F(x)\le 1 \Rightarrow 0\le a\le 1$ và $b>0$

2.17 $\mathrm{E}(X)=-0,3$; $\mathrm{D}(X)=15,21$.

2.18 $x_3=29,1$; $p_3=0,2$.

2.19 $\mathrm{E}(X_1)=3,1$; $\mathrm{E}(X_2)=3,4$; $\mathrm{D}(X_1)=1,09$; $\mathrm{D}(X_2)=1,44$.

$\mathrm{E}(X_1+X_2)=6,5$; $\mathrm{D}(X_1+X_2)=2,53$.

2.20 $\mathrm{E}(\overline{X})=0,8$; $\mathrm{D}(\overline{X})=0,12$.

2.21 a) $\mathrm{D}(Z)=61$. b) $\mathrm{D}(Z)=41$.

2.22 $p_1=0,4$; $p_2=0,1$; $p_3=0,5$.

2.23 a) Gọi A_i là biến cố toa i có người ngồi ($i=\overline{1,3}$). Gọi A là biến cố cả 3 toa đều có người ngồi. Khi đó: $A=A_1A_2A_3$

$\Rightarrow \overline{A}=\overline{A_1}+\overline{A_2}+\overline{A_3}$.

$$P(\overline{A})=P(\overline{A}_1)+P(\overline{A}_2)+P(\overline{A}_3)-P(\overline{A}_1\overline{A}_2)-P(\overline{A}_1\overline{A}_3)$$
$$-P(\overline{A}_2\overline{A}_3)+P(\overline{A}_1\overline{A}_2\overline{A}_3)=\frac{31}{81}$$

$\Rightarrow P(A)=\dfrac{50}{81}$.

b)

X	0	1	2	3	4	5
P	$\frac{32}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{40}{243}$	$\frac{10}{243}$	$\frac{1}{243}$

Y	0	1	2	3	4	5
P	$\frac{1}{243}$	$\frac{10}{243}$	$\frac{40}{243}$	$\frac{80}{243}$	$\frac{80}{243}$	$\frac{32}{243}$

2.24 $\mathrm{E}(X)=0$.

2.25 a) Vì $\displaystyle\int_0^4 x^2(4-x)\,dx=\frac{64}{3} \Rightarrow k=\frac{3}{64}$.

$$P\{X<1\}=\int_0^1 \frac{3}{64}x^2(4-x)\,dx=\frac{13}{256}.$$

b) $\displaystyle \mathrm{E}X=\int_0^4 \frac{3}{64}x^3(4-x)\,dx=\frac{3}{64}\left(x^4-\frac{x^5}{5}\right)\Bigg|_{x=0}^{x=4}=3\left(4-\frac{16}{5}\right)=\frac{12}{5}$

$$\mathrm{E}X^2 = \int_0^4 \frac{3}{64}x^4(4-x)\,dx = \frac{3}{64}\left(\frac{4x^5}{5} - \frac{x^6}{6}\right)\Bigg|_{x=0}^{x=4} = 3\times 64\left(\frac{1}{5} - \frac{1}{6}\right) = \frac{32}{5}$$

$$\Rightarrow \mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = \frac{32}{5} - \left(\frac{12}{5}\right)^2 = \frac{16}{25}.$$

2.26 a) Kí hiệu A_i là biến cố: "A bắn trúng i viên",

B_i là biến cố: "B bắn trúng i viên"; i = 0, 1 2. Dễ thấy:

$P(A_0) = 0,36$; $P(A_1) = 0,48$; $P(A_2) = 0,16$;

$P(B_0) = 0,25$; $P(B_1) = 0,5$; $P(B_2) = 0,25.$

Từ đó $P\{X=-2\} = P(A_0)P(B_2) = 0,09$

$P\{X=-1\} = P(A_0)P(B_1) + P(A_1)P(B_2) = 0,18 + 0,12 = 0,3$

$P\{X=0\} = P(A_0)P(B_0) + P(A_1)P(B_1) + P(A_2)P(B_2) = 0,37$

$P\{X=1\} = P(A_1)P(B_0) + P(A_2)P(B_1) = 0,2$

$P\{X=2\} = P(A_2)P(B_0) = 0,04$

Vậy bảng phân bố xác suất của X

X	-2	-1	0	1	2
P	0,09	0,3	0,37	0,2	0,04

$$\mathrm{E}X = (-2)\times 0,09 + (-1)\times 0,3 + 0\times 0,37 + 1\times 0,2 + 2\times 0,04 = -0,2$$

$$\mathrm{E}X^2 = (-2)^2\times 0,09 + (-1)^2\times 0,3 + 0^2\times 0,37 + 1^2\times 0,2 + 2^2\times 0,04 = 1,02$$

$$\mathrm{D}X = \mathrm{E}X^2 - (\mathrm{E}X)^2 = 1,02 - (-0,2)^2 = 0,98$$

b) $P\{Y=0\} = 0,37$

$P\{Y=1\}=P\{X=1\}+P\{X=-1\}=0,5$

$P\{Y=2\}=P\{X=2\}+P\{X=-2\}=0,13$

$EY=0\times0,37+1\times0,5+2\times0,13=0,76$.

2.27 Kí hiệu A_i là biến cố: "ô tô thứ i bị hỏng", i = 1, 2.

Dễ thấy $P(A_1)=0,1;\ P(A_2)=0,2$

Gọi X là số ôtô bị hỏng trong thời gian làm việc

Từ đó $P\{X=0\}=P(\overline{A_1})P(\overline{A_2})=0,9\times0,8=0,72$

$P\{X=1\}=P(A_1)P(\overline{A_2})+P(\overline{A_1})P(A_2)=0,1\times0,8+0,9\times0,2=0,26$

$P\{X=2\}=P(A_1)P(A_2)=0,02$

X	0	1	2
P	0,72	0,26	0,02

$EX=0\times0,72+1\times0,26+2\times0,02=0,3$.

$EX^2=0^2\times0,72+1^2\times0,26+2^2\times0,02=0,34$.

$DX=EX^2-(EX)^2=0,34-(0,3)^2=0,25$.

2.28 a) Điều kiện $\begin{cases}p_i>0\\ \sum_i p_i=1\end{cases}\Rightarrow\begin{cases}k>0\\ 1\ 0^2+\ k9=\end{cases}\Rightarrow\begin{cases}k>0\\ \left[\begin{array}{l}k=-1\\ k=1\ /\end{array}\right.\end{cases}$

$\Rightarrow k=1/10$.

b) $P\{X\ge5\}=\frac{1}{100}+\frac{2}{100}+\frac{7}{100}+\frac{1}{10}=\frac{2}{10}$; $P\{X<3\}=\frac{3}{10}$.

c) $EX=\frac{1}{10}+\frac{4}{10}+\frac{6}{10}+\frac{12}{10}+\frac{5}{100}+\frac{12}{100}+7\left(\frac{7}{100}+\frac{1}{10}\right)=3,66$.

d) $\mathrm{E}X^2 = \frac{1}{10} + \frac{8}{10} + \frac{18}{10} + \frac{48}{10} + \frac{25}{100} + \frac{72}{100} + 49\left(\frac{7}{100} + \frac{1}{10}\right) = 16,8$.

$\mathrm{D}X = \mathrm{E}X^2 - \left(\mathrm{E}X\right)^2 = 16,8 - \left(3,66\right)^2 = 3,404$.

2.29 a) Gọi X là "số phế phẩm gặp phải":

X	0	1
P	0,6	0,4

$\mathrm{E}X = \frac{2}{5}$; $\mathrm{D}X = \frac{6}{25}$.

b) Gọi Y là "số chính phẩm gặp phải" $\Rightarrow Y = 2 - X$:

Y	0	1
P	0,4	0,6

$\mathrm{E}Y = \mathrm{E}\left(2 - X\right) = 2 - \frac{2}{5} = \frac{8}{5}$; $\mathrm{D}Y = \frac{6}{25}$.

2.30 Gọi X là "số nữ có trong nhóm được chọn"

$$P\{X=0\} = \frac{C_6^3}{C_{10}^3} = \frac{5}{30}\ ,\quad P\{X=1\} = \frac{C_4^1 C_6^2}{C_{10}^3} = \frac{15}{30}$$

$$P\{X=2\} = \frac{C_4^2 C_6^1}{C_{10}^3} = \frac{9}{30}\ ,\quad P\{X=1\} = \frac{C_4^1 C_6^2}{C_{10}^3} = \frac{15}{30}$$

X	0	1	2	3
P	5/30	15/30	9/30	1/30

$\mathrm{E}X = 0\times\frac{5}{30} + 1\times\frac{15}{30} + 2\times\frac{9}{30} + 3\times\frac{1}{30} = \frac{36}{30} = \frac{6}{5}$.

2.31 Thắng 2 trong 4 ván dễ hơn.

2.32 a) $P = 0{,}238$. b) $P = 0{,}751$.

2.33 a) X tuân theo quy luật nhị thức $\mathcal{B}(n;p)$ với $n = 5$ và $p = 0{,}8$.

b) $\mathrm{E}X = 4;\ \mathrm{D}X = 0{,}8$. c) $\mathrm{Mod}X = 4$; $P\{X = 4\} = 0{,}4096$.

2.34 a) $P = 0{,}9914$

b) Số sản phẩm hỏng trung bình là 0,5

c) Số sản phẩm hỏng có khả năng xảy ra nhiều nhất là 0.

2.35 Gọi x là số câu hỏi học sinh trả lời đúng. Số điểm anh ta nhận được là:

$$4x + (10 - x)(-2) = 6x - 20$$

a) Anh ta được 4 điểm khi trả lời đúng: $6x - 20 = 4 \Rightarrow x = 4$.

Vậy xác suất để anh ta được điểm 4 là:

$$P = C_{10}^{4}\left(\frac{1}{5}\right)^{4}\left(\frac{4}{5}\right)^{6} = 0{,}088$$

b) Anh ta được điểm âm khi trả lời đúng:

$6x - 20 < 0 \Rightarrow x = 0,1,2,3$

Vậy xác suất để anh ta được điểm âm là:

$$P = \sum_{k=0}^{3} C_{10}^{4}\left(\frac{1}{5}\right)^{k}\left(\frac{4}{5}\right)^{10-k} = 0{,}879$$

2.36 Gọi X là biến ngẫu nhiên đếm số các chữ số 1 trong dãy có năm chữ số. Vì chỉ có hai khả năng xuất hiện là chữ số 1 và 0, xác suất sinh ra chữ số 1 không thay đổi vì vậy X có phân bố nhị thức tham số $(n;p) = (5;0,6)$.

a) $P\{X = 2\} = C_5^2(0,6)^2(0,4)^3 = 0,23$

b) $P\{X \le 2\} = \sum_{k=0}^{2} C_5^k (0,6)^k (0,4)^{5-k} = 0,317$

$\Rightarrow P\{X \ge 3\} = 1 - P\{X \le 2\} = 1 - 0,317 = 0,683$.

2.37 Gọi X là số lần thu được tín hiệu trong 5 lần phát độc lập thì $X \sim B(5;\ 0,7)$

a) Xác suất thu được tín hiệu 2 lần:

$P\{X = 2\} = C_5^2 0,7^2 0,3^3 = 0,132$

b) Xác suất thu được tín hiệu nhiều nhất 1 lần: $P\{X \le 1\} = 0,031$

c) Xác suất thu được tín hiệu:

$P\{X \ge 1\} = 1 - P\{X = 0\} = 1 - 0,002 = 0,998$

2.38 Không đúng; $P = 0,41$.

2.39 a) Gọi X là số cuộc gọi trong khoảng thời gian 10 giây thì X có phân bố Poisson tham số $\lambda = 1/3$. Vậy xác suất có ít nhất một cuộc gọi trong khoảng thời gian 10 giây là:

$P\{X \ge 1\} = 1 - P\{X = 0\} = 1 - e^{-1/3} = 0,2835$.

b) Gọi Y là số cuộc gọi trong khoảng thời gian 3 phút thì Y có phân bố Poisson tham số $\lambda = 6$. Vậy xác suất có nhiều nhất ba cuộc gọi trong khoảng thời gian 3 phút là $P\{Y \le 3\} = 0,151$.

c) Gọi Z là số cuộc gọi trong khoảng thời gian 1 phút thì Z có phân bố Poisson tham số $\lambda = 2$. Xác suất có nhiều nhất 1 cuộc gọi trong khoảng thời gian 1 phút là $P\{Z \le 1\} = 0,406$. Vậy xác suất để trong khoảng thời gian 3 phút liên tiếp mỗi phút có nhiều nhất 1 cuộc gọi là:

$P\{Z \le 1\}^3 = 0,406^3 = 0,0067$.

2.40 a) $\sum_{k=1}^{\infty} p_X(k) = \sum_{k=1}^{\infty} q^{k-1}p = p\sum_{k=0}^{\infty} q^k = \frac{p}{1-q} = 1$

$P\{X > k\} = \sum_{i=k+1}^{\infty} q^{i-1}p = \frac{q^k p}{1-q} = q^k$.

Vậy $P\{X \le k\} = 1 - P\{X > k\} = 1 - q^k$

Do đó $F_X(x) = P\{X \le x\} = 1 - q^k$ nếu $k \le x < k+1$.

b) Xét hàm số $g(x) = \sum_{n=0}^{\infty} ax^n = \frac{a}{1-x}$ với $|x| < 1$ thì:

$\sum_{n=1}^{\infty} anx^{n-1} = g'(x) = \frac{a}{(1-x)^2}$, $\sum_{n=2}^{\infty} an(n-1)x^{n-2} = g''(x) = \frac{2a}{(1-x)^3}$.

Áp dụng kết quả này ta được $\mathrm{E}(X) = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$

$\mathrm{E}[X(X-1)] = \frac{2pq}{(1-q)^3} = \frac{2pq}{p^3} = \frac{2(1-p)}{p^2}$

$\Rightarrow \mathrm{E}[X^2] = \mathrm{E}[X(X-1)] + \mathrm{E}(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$

$\mathrm{D}(X) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2} = \frac{q}{p^2}$.

2.41 $P\{8 < X < 12\} = \Phi\left(\frac{12-10}{2}\right) - \Phi\left(\frac{8-10}{2}\right) = 0{,}6826$.

2.42 $P = 0{,}3$.

2.43 a) 95,44%; b) 4,56%.

2.44 a) 20,33%; b) $P = 0{,}9983$.

2.45 $\mathrm{E}(\overline{X}) = \mu;\ \mathrm{D}(\overline{X}) = \dfrac{\sigma^2}{n} \ \Rightarrow\ P\left\{\left|\overline{X} - \mu\right| < \varepsilon\right\} = 2\Phi\left(\dfrac{\varepsilon\sqrt{n}}{\sigma}\right) - 1.$

2.46 a) Hàm phân bố $F_X(x) = \int\limits_0^x \dfrac{t}{\sigma^2} e^{-t^2/(2\sigma^2)} dt\ ,\ x \geq 0$

Đặt $y = \dfrac{t^2}{2\sigma^2}$ thì $dy = \dfrac{t}{\sigma^2} dt$ và $F_X(x) = \int\limits_0^{x^2/(2\sigma^2)} e^{-y} dy = 1 - e^{-x^2/(2\sigma^2)}$.

b) $\mathrm{E}\,X = \int\limits_0^\infty \dfrac{x^2}{\sigma^2} e^{-x^2/(2\sigma^2)} dx = \sigma \int\limits_0^\infty y^2 e^{-y^2/2} dy = \dfrac{\sigma}{2}.$

$$\mathrm{E}\,X^2 = \int\limits_0^\infty \frac{x^3}{\sigma^2} e^{-x^2/(2\sigma^2)} dx = \int\limits_0^\infty 2\sigma^2 y e^{-y} dy = 2\sigma^2.$$

$$\mathrm{D}\,X = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 = \frac{7\sigma^2}{4}.$$

Đáp án chương 3

Câu hỏi	**3.1**	**3.2**	**3.3**	**3.4**	**3.5**	**3.6**	**3.7**	**3.8**
Đáp án	Sai	Đúng	Đúng	Đúng	Sai	Sai	Sai	Đúng
Câu hỏi	**3.9**	**3.10**	**3.11**	**3.12**				
Đáp án	Đúng	Sai	Sai	Đúng				

3.13

X	x_1	x_2
P	0,56	0,44

Y	y_1	y_2	y_3
P	0,26	0,38	0,36

3.14 $\mathrm{E}X = -1/8;\ \ \mathrm{E}Y = 0;$

$\mathrm{cov}(X, Y) = -1/8;\ \ \rho_{X,Y} = -0{,}15.$

3.15 $EX = -1/5$; $EY = 0$; $\rho_{X,Y} = 0$. X và Y không độc lập vì

$P\{X = 1\} = 2/15$, $P\{Y = 1\} = 5/15$ và $P\{X = 1.\ Y = 1\} = 0$.

3.16 Bảng phân bố xác suất của biến ngẫu nhiên Z

Z	1	2	3	4	6
P	0,12	0,43	0,03	0,35	0,07

$EX = 1{,}7$; $EY = 1{,}7$; $EZ = 2{,}89$.

3.17 Bảng phân bố xác suất đồng thời của X và Y

X \ Y	0	1	2	3	4
0	0,04	0,12	0,16	0,06	0,02
1	0,03	0,09	0,12	0,045	0,015
2	0,02	0,06	0,08	0,03	0,01
3	0,01	0,03	0,04	0,015	0,005

$P\{X > Y\} = 0{,}19$.

3.18 X, Y không độc lập vì

$P\{X = 1\} = 0{,}5$, $P\{Y = 1\} = 0{,}45$ và $P\{X = 1.\ Y = 1\} = 0{,}15 \neq 0{,}5 \cdot 0{,}45$

$$P\{X = 1 | Y = 2\} = \frac{P\{X = 1; Y = 2\}}{P\{Y = 2\}} = \frac{0{,}2}{0{,}25} = 0{,}8.$$

3.19

X \ Y	1	2	3	4	5	6
0	1/12	1/12	1/12	1/12	1/12	1/12
1	1/12	1/12	1/12	1/12	1/12	1/12

$P\{X = 1\} = 0{,}5$, $P\{Y = 1\} = 0{,}45$ và $P\{X = 1.\ Y = 1\} = 0{,}15 \neq 0{,}5 \cdot 0{,}45$

3.20

Y/X = 26	23	27
P	0,357	0,643

X/Y = 27	26	30	41	50
P	0,1268	0,4225	0,1549	0,2958

3.21 $\mathrm{E}\left[Y \middle| X=1\right]=5$. $\mathrm{E}X=2{,}93$; $\mathrm{E}Y=4{,}5$; $\mathrm{D}X=4{,}83$; $\mathrm{D}Y=2{,}25$.

3.22 a. $\alpha=15$; $\mathrm{E}X=-0{,}2$; $\mathrm{E}Y=0$.

b. $\mathrm{cov}(X,Y)=0 \Rightarrow \rho(X,Y)=0$.

c. X, Y không độc lập vì $P\{X=1\}=\dfrac{2}{15}$, $P\{Y=1\}=\dfrac{5}{15}$ nhưng $P\{X=1,Y=1\}=0$.

3.23 Hàm khối lượng xác suất của biến ngẫu nhiên rời rạc X là

$$p_X(k)=e^{-\lambda}\frac{\lambda^k}{k!}.$$

$$P(B)=P\{X=0,2,4,...\}=\sum_{k=0}^{\infty}e^{-\lambda}\frac{\lambda^{2k}}{(2k)!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!}$$

$$e^{\lambda}=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=\sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!}+\sum_{k=0}^{\infty}\frac{\lambda^{2k+1}}{(2k+1)!}$$

$$e^{-\lambda}=\sum_{k=0}^{\infty}\frac{(-\lambda)^k}{k!}=\sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!}-\sum_{k=0}^{\infty}\frac{\lambda^{2k+1}}{(2k+1)!}$$

$$\sum_{k=0}^{\infty}\frac{\lambda^{2k}}{(2k)!}=\frac{e^{\lambda}+e^{-\lambda}}{2} \Rightarrow P(B)=e^{-\lambda}\frac{e^{\lambda}+e^{-\lambda}}{2}=\frac{1+e^{-2\lambda}}{2}$$

$$p_X(k\mid B)=\frac{P\{(X=k)B\}}{P(B)};\ \{(X=2k)B\}=\{X=2k\},$$

$\{(X = 2k+1)B\} = \varnothing$.

$$p_X(k \mid B) = \begin{cases} \dfrac{P\{X=k\}}{P(B)} = \dfrac{2e^{-\lambda}\lambda^k}{(1+e^{-2\lambda})k!} & k \text{ chẵn} \\ \dfrac{\varnothing}{P(B)} = 0 & k \text{ lẻ} \end{cases}$$

3.24 a) $k = 3$;

$$f_X(x) = \begin{cases} 3x^2 & \text{nếu } 0 < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases};$$

$$f_Y(y) = \begin{cases} \dfrac{3}{2}(1-y^2) & \text{nếu } 0 < y < 1 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

X và Y không độc lập vì $P\left\{X < \dfrac{1}{2}, Y > \dfrac{1}{2}\right\} = 0$ nhưng $P\left\{X < \dfrac{1}{2}\right\} \neq 0$, $P\left\{Y > \dfrac{1}{2}\right\} \neq 0$.

3.25 Áp dụng công thức (3.14) ta được

$$f(x,y) = \frac{\partial^2 F}{\partial x \partial y} = \begin{cases} e^{-x-y} & \text{nếu } x > 0, y > 0; \\ 0 & \text{nếu ngược lại}. \end{cases}$$

Áp dụng công thức (3.53) ta được

$$f(x|y) = \begin{cases} e^{-x} & \text{nếu } x > 0, \\ 0 & \text{nếu } x \leq 0. \end{cases}$$

3.26 a) $k = \dfrac{1}{\pi^2}$;

b) $F(x,y) = \left(\dfrac{1}{\pi}\operatorname{arctg} x + \dfrac{1}{2}\right)\left(\dfrac{1}{\pi}\operatorname{arctg} y + \dfrac{1}{2}\right)$;

$$F_X(x) = \lim_{y\to\infty} F(x,y) = \frac{1}{\pi}\operatorname{arctg} x + \frac{1}{2};$$

$$F_Y(y) = \lim_{x\to\infty} F(x,y) = \frac{1}{\pi}\operatorname{arctg} y + \frac{1}{2};$$

Vì $F(x,y) = F_X(x)F_Y(y)$ nên ta kết luận X và Y độc lập.

d) $P\left\{1 < X < \sqrt{3}, 0 < Y < 1\right\}$

$$= P\left\{1 < X < \sqrt{3}\right\} P\left\{0 < Y < 1\right\} = \frac{1}{12}\cdot\frac{1}{4} = \frac{1}{48}$$

3.27 $F_X(x \mid a < X \le b) = P\left\{X \le x \mid a < X \le b\right\} = \dfrac{P\left\{X \le x; a < X \le b\right\}}{P\left\{a < X \le b\right\}}$

$$\left\{X \le x; a < X \le b\right\} = \begin{cases} \varnothing & x \le a \\ \left\{a < X \le x\right\} & a < x \le b \\ \left\{a < X \le b\right\} & x > b \end{cases}$$

$$\Rightarrow F_X(x \mid a < X \le b) = \begin{cases} 0 & x \le a \\ \dfrac{F_X(x) - F_X(a)}{F_X(b) - F_X(a)} & a < x \le b \\ 1 & x > b \end{cases}$$

3.28 Tương tự bài tập 3.27 ta có:

$$F_X(x \mid X > 0) = \begin{cases} 0 & x \le 0 \\ 2F_X(x) - 1 & x > 0 \end{cases}$$

$$f_X(x \mid X > 0) = \begin{cases} 0 & x < 0 \\ \dfrac{2}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)} & x \ge 0 \end{cases};$$

$$\mathrm{E}\left[X \mid X > 0\right] = \frac{2}{\sigma\sqrt{2\pi}} \int_0^\infty x e^{-x^2/(2\sigma^2)} = \sigma\sqrt{\frac{2}{\pi}}$$

3.29 Hàm mật độ của X là $f_X(x) = \begin{cases} \dfrac{\ln x}{x^2} & \text{nếu } x \geq 1, \\ 0 & \text{nếu } x < 1. \end{cases}$

Hàm mật độ của Y là $f_Y(y) = \begin{cases} \dfrac{1}{2y^2} & \text{nếu } 1 \leq y < \infty, \\ \dfrac{1}{2} & \text{nếu } 0 \leq y \leq 1. \end{cases}$

Từ đó hàm mật độ có điều kiện của Y với điều kiện $X = x$ $(x > 1)$ là : $f(y|x) = \dfrac{f(x,y)}{f_X(x)} = \dfrac{1}{2y \ln x}$ nếu $\dfrac{1}{x} \leq y \leq x$;

Hàm mật độ có điều kiện của X với điều kiện $Y = y$ $(y > 0)$ là :

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} = \begin{cases} \dfrac{1}{x^2 y} & \text{nếu } 0 \leq y \leq 1, x \geq y \\ \dfrac{y}{x^2} & \text{nếu } y \geq 1, y \leq x. \end{cases}$$

3.30 a) $k = \dfrac{1}{8}$; $f_X(x) = \begin{cases} \dfrac{1}{4}(x+1) & \text{nếu } 0 < x < 2 \\ 0 & \text{nếu ngược lại} \end{cases}$,

$$f_Y(y) = \begin{cases} \dfrac{1}{4}(y+1) & \text{nếu } 0 < y < 2 \\ 0 & \text{nếu ngược lại} \end{cases}$$

X , Y không độc lập.

b) $f_{X|Y}(x|y) = \dfrac{1}{2}\dfrac{x+y}{y+1}$ $0 < x < 2, 0 < y < 2$;

$f_{Y|X}(y|x) = \dfrac{1}{2}\dfrac{x+y}{x+1}$ $0 < x < 2, 0 < y < 2$.

c) $P\{0<Y<1/2 \mid X=1\} = \int_0^{1/2} f_{Y|X}(y \mid x=1)dy$.

$$= \frac{1}{2}\int_0^{1/2}\left(\frac{1+y}{2}\right)dy = \frac{5}{32}$$

3.31 $\mathrm{E}(2X-3Y) = 2\mathrm{E}(X) - 3\mathrm{E}(Y) = 10$;

$\mathrm{D}(2X-3Y) = 4\mathrm{D}(X) + 9\mathrm{D}(Y) - 12\sqrt{\mathrm{D}(X)\mathrm{D}(Y)}\,\rho_{X,Y} = 57{,}6$.

3.32 Xét $F(t) = \mathrm{E}\left[X - tY\right]^2 \geq 0,\ \forall t$

$$F(t) = \mathrm{E}\left[X\right]^2 - 2t\,\mathrm{E}\left[XY\right] + t^2\,\mathrm{E}\left[Y\right]^2$$

$$\Rightarrow \Delta' = \left(\mathrm{E}[XY]\right)^2 - \left(\mathrm{E}\,X^2\right)\left(\mathrm{E}\,Y^2\right) \leq 0.$$

3.33 $f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)$.

$$a>0:\ F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt = \frac{1}{a}\int_{-\infty}^{y} f_X\left(\frac{t-b}{a}\right)dt = F_X\left(\frac{y-b}{a}\right).$$

$$a<0:\ F_Y(y) = \int_{-\infty}^{y} f_Y(t)dt = -\frac{1}{a}\int_{-\infty}^{y} f_X\left(\frac{t-b}{a}\right)dt = 1 - F_X\left(\frac{y-b}{a}\right).$$

Đáp án chương 4

Câu hỏi	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8
Đáp án	Đúng	Đúng	Sai	Sai	Đúng	Đúng	Đúng	Đúng

4.9 Gọi X là số máy hỏng trong ca. X có phân bố nhị thức $\mathrm{E}X = 0{,}5$, $\mathrm{D}X = 1$. Áp dụng bất đẳng thức Trêbusép ta có:

$$P\{|X-0{,}05|<2\}\geq 1-\frac{0{,}475}{2^2}=0{,}88;$$

$$P\{|X-0{,}05|\geq 2\}\leq \frac{0{,}475}{2^2}=0{,}12.$$

4.10 Đặt $S=\sum_{n=1}^{12}X_n$; $\mathrm{E}S=12\cdot 16=192$, $\mathrm{D}S=12$. Theo bất đẳng thức Trêbusép:

$$P\{|S-192|\leq \varepsilon\}\geq 1-\frac{\mathrm{D}S}{\varepsilon^2}\geq 0{,}99.$$ Chọn $a=157{,}36$; $b=226{,}64$.

4.11 Đặt $S=\sum_{n=1}^{10000}X_n$; $\mathrm{E}S=0$, $\mathrm{D}S=\frac{10000}{12}$. Theo bất đẳng thức Trêbusép:

$$P\{|S|\geq 500\}\leq \frac{\mathrm{D}S}{500^2}=\frac{1}{300}.$$

4.12 Áp dụng công thức (4.6) cho biến ngẫu nhiên $\overline{X}$ ta được.

$$P\left\{\left|\overline{X}-\mu\right|\leq\frac{\sigma}{10}\right\}\geq 1-\frac{\sigma^2/n}{\sigma^2/100}=1-\frac{100}{n}.$$

Chọn $\frac{100}{n}\leq 0{,}05$ hoặc $n\geq 100/0{,}5=200$ thì:

$$P\left\{\left|\overline{X}-\mu\right|\leq\frac{\sigma}{10}\right\}\geq 0{,}95.$$

4.13 Ta biết rằng S là biến ngẫu nhiên có phân bố nhị thức tham số $p=\frac{1}{6}$. $\mathrm{E}S=\frac{n}{6}$ và $\mathrm{D}S=\frac{5n}{36}$. Theo bất đẳng thức Trêbusép:

$$P\{|S-\mathrm{E}S|<\sqrt{n}\}\geq 1-\frac{\mathrm{D}S}{n}=1-\frac{5}{36}=\frac{31}{36}\Leftrightarrow P\left\{\frac{n}{6}-\sqrt{n}<S<\frac{n}{6}+\sqrt{n}\right\}\geq\frac{31}{36}.$$

4.14 Đặt $S = \sum_{n=1}^{12} X_n$. Ta cần tìm M nhỏ nhất để:

$$P\left\{\sum_{n=1}^{12} X_n \le M\right\} \ge 0{,}99 .$$

Ta có $\mathrm{E}S = 192$, $\mathrm{D}S = 12$. Theo bất đẳng thức Trêbusép:

$$P\{|S - 192| \le \varepsilon\} \ge 1 - \frac{\mathrm{D}S}{\varepsilon^2} \ge 0{,}99 \Rightarrow \varepsilon = 34{,}64 .$$

Vậy M = 192+34,64 = 226,64.

4.15 Thỏa mãn luật số lớn Trêbusép.

4.16 Thỏa mãn luật số lớn Trêbusép.

4.17 Thỏa mãn luật số lớn Trêbusép.

4.18 Áp dụng bất đẳng thức Trêbusép tính được xác suất $P \ge 0{,}9131$

4.19 Áp dụng bất đẳng thức Trêbusép cần kiểm tra 23.750 chi tiết.

4.20 Gọi X là số sản phẩm hỏng. Ta có $X \sim \mathcal{B}(250; 0{,}02)$. X sẽ có xấp xỉ phân bố Poisson với $\lambda = 250 \cdot 0{,}02 = 5$. Từ đó tra bảng ta được:

a) $P\{X = 2\} = 0{,}0842$;

b) $P\{X \le 2\} = 0{,}1247$

4.21 Giả sử X là số người chọn ăn ở đợt 1. Khi đó $1000 - X$ là số người chọn ăn ở đợt 2 . Gọi k là số chỗ ngồi trong nhà ăn. Ta phải chọn k nhỏ nhất để:

$$P\{X < k,\ 1000 - X < k\} \ge 0{,}99 \Leftrightarrow P\{1000 - k < X < k\} \ge 0{,}99$$

Ta xem X có phân bố chuẩn với $\mu = 500$, $\sigma = \sqrt{250}$.

Vậy ta có:

$$\Phi\left(\frac{k-500}{\sqrt{250}}\right)-\Phi\left(\frac{500-k}{\sqrt{250}}\right)\geq 0,99 \Leftrightarrow 2\Phi\left(\frac{k-500}{\sqrt{250}}\right)\geq 1,99$$

$$\Leftrightarrow \Phi\left(\frac{k-500}{\sqrt{250}}\right)\geq \Phi(2,58)$$

Từ đó $k \geq 500 + 2,58\sqrt{250} = 540,49$. Vậy $k = 541$.

4.22 a) Gọi X là số người trúng tuyển. Ta có $X \sim \mathcal{B}(325;0,9)$. X có phân bố xấp xỉ chuẩn với $\mu = 292,5$, $\sigma^2 = 29,25$. Vậy $P\{X \leq 300\} \approx \Phi\left(\frac{300-292,5}{\sqrt{29,25}}\right) = \Phi(1,3868) \approx 0,917$.

b) Giả sử n là số người dự thi. Phân bố của X xấp xỉ phân bố chuẩn với $\mu = 0,9n$, $\sigma = 0,3\sqrt{n}$. Vậy:

$$P\{X \leq 300\} \approx \Phi\left(\frac{300-0,9n}{0,3\sqrt{n}}\right) \geq 0,99 = \Phi(2,33)$$

$$\Leftrightarrow 300 - 0,9n \geq (0,3)(2,33)\sqrt{n}$$

Giải bất phương trình ta được $n \leq 319,51$. Vậy $n = 319$.

4.23 a) Rõ ràng rằng số lỗi trong 10 ký số nhận được là biến ngẫu nhiên X có phân bố nhị thức với tham số $(n;p) = (10;0,01)$

$$P\{X>1\} = 1 - P\{X=0\} - P\{X=1\}$$
$$= 1 - C_{10}^0(0,01)^0(0,99)^{10} - C_{10}^1(0,01)^1(0,99)^9 = 0,0042$$

b) Sử dụng công thức (4.16) với $\lambda = np = 10(0,01) = 0,1$

$$P\{X>1\} = 1 - P\{X=0\} - P\{X=1\}$$
$$= 1 - e^{-0,1}\frac{(0,1)^0}{0!} - e^{-0,1}\frac{(0,1)^1}{1!} = 0,0047$$

4.24 Theo công thức (2.24) ta có thể xem $X = X_1 + \cdots + X_n$, trong đó $X_1,...,X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố Poisson tham số $\frac{\lambda}{n}$. Kỳ vọng và phương sai $\mathrm{E}X = \mathrm{D}X = \lambda$.

Áp dụng định lý giới hạn trung tâm, công thức (4.11) ta được

$$P\left\{\frac{X-\lambda}{\sqrt{\lambda}} \le y\right\} \approx \Phi(y)$$

Mặt khác $P\left\{\frac{X-\lambda}{\sqrt{\lambda}} \le y\right\} = P\left\{X \le \sqrt{\lambda}y + \lambda\right\}$

$$\Rightarrow P\{X \le x\} \approx \Phi\left(\frac{x-\lambda}{\sqrt{\lambda}}\right).$$

Đáp án chương 5

Câu hỏi	**5.1**	**5.2**	**5.3**	**5.4**	**5.5**	**5.6**	**5.7**	**5.8**	**5.9**
Đáp án	Đúng	Sai	Đúng	Đúng	Đúng	Sai	Sai	Đúng	Đúng
Câu hỏi	**5.10**	**5.11**	**5.12**	**5.13**	**5.14**	**5.15**	**5.16**	**5.17**	**5.18**
Đáp án	Sai	Sai	Sai	Đúng	Đúng	Sai	Sai	Sai	Sai

5.19 Mẫu ngẫu nhiên có kích thước 10: $W = (X_1, X_2, ..., X_{10})$.

$$P\left\{\overline{X} = \frac{1}{2}\right\} = P\left\{\frac{1}{10}\sum_{i=1}^{10} X_i = \frac{1}{2}\right\} = P\left\{\sum_{i=1}^{10} X_i = 5\right\}.$$

Vì X có phân bố nhị thức nên

$$P\left\{\sum_{i=1}^{10} X_i = 5\right\} = P_{10}(5) = C_{10}^5 (0{,}5)^5 \cdot (0{,}5)^{10-5} = C_{10}^5 (0{,}5)^{10}.$$

5.20 X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ nên $\overline{X}$ có phân bố chuẩn $\mathbf{N}(\mu;\frac{\sigma^2}{n})$

Vậy $P\left\{\left|\overline{X} - \mu\right| < \varepsilon\right\} = P\left\{\mu - \varepsilon < \overline{X} < \mu + \varepsilon\right\}$

$$= \Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(\frac{-\varepsilon\sqrt{n}}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - 1 .$$

Do đó: $P\left\{\left|\overline{X} - 20\right| < 0,2\right\} = 2\Phi\left(\frac{0,2\sqrt{100}}{1}\right) - 1 = 2\Phi(2) - 1 = 0,9545 .$

5.21 Bảng phân bố tần số

X	1	2	3	4
Tần số	2	4	2	2

Bảng phân bố tần suất

X	1	2	3	4
Tần suất	1/5	2/5	1/5	1/5

Hàm phân bố thực nghiệm

$$F_{10}(x) = \begin{cases} 0 & x \le 1 \\ 1/5 & 1 < x \le 2 \\ 3/5 & 2 < x \le 3 \\ 4/5 & 3 < x \le 4 \\ 1 & x > 4 \end{cases}$$

$\overline{x} = 2,4 ;\ s^2 = 1,15 ,\ \ s = 1,072 .$

5.22 Áp dụng bất đẳng thức Trêbusép ta có:

$$P\left\{\left|\Theta_n - \theta\right| \ge \varepsilon\right\} \le \frac{\mathrm{E}\left[(\Theta_n - \theta)^2\right]}{\varepsilon^2} = \frac{1}{\varepsilon^2}\mathrm{E}\left[\left\{\left(\Theta_n - \mathrm{E}\,\Theta_n\right) + \left(\mathrm{E}\,\Theta_n - \theta\right)\right\}^2\right]$$

$$= \frac{1}{\varepsilon^2}\mathrm{E}\left[\left(\Theta_n - \mathrm{E}\,\Theta_n\right)^2 + \left(\mathrm{E}\,\Theta_n - \theta\right)^2 + 2\left(\Theta_n - \mathrm{E}\,\Theta_n\right)\left(\mathrm{E}\,\Theta_n - \theta\right)\right]$$

$$= \frac{1}{\varepsilon^2}\left[\mathrm{D}\left(\Theta_n\right) + \mathrm{E}\left(\mathrm{E}\,\Theta_n - \theta\right)^2 + 2\,\mathrm{E}\left(\Theta_n - \mathrm{E}\,\Theta_n\right)\left(\mathrm{E}\,\Theta_n - \theta\right)\right].$$

Vì vậy nếu $\lim\limits_{n\to\infty} \mathrm{E}\left(\Theta_n\right) = \theta$ và $\lim\limits_{n\to\infty} \mathrm{D}\left(\Theta_n\right) = 0$ thì:

$$\lim_{n\to\infty} P\left\{\left|\Theta_n - \theta\right| \geq \varepsilon\right\} = 0 .$$

5.23 Hàm hợp lý của tham số λ

$$L(\lambda) = f(x_1,...,x_n;\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{x_1!...x_n!}$$

$$\Rightarrow \ln L(\lambda) = -n\lambda + \ln\lambda \sum_{i=1}^{n} x_i - \ln(x_1!...x_n!)$$

$$\Rightarrow \frac{d}{d\lambda}\ln L(\lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i = 0 \Rightarrow \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x} .$$

Vậy ước lượng hợp lý cực đại của tham số λ là $\Lambda = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i = \overline{X}$.

5.24 Hàm hợp lý của tham số λ

$$L(\lambda) = f(x_1,...,x_n;\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda\sum_{i=1}^{n} x_i}$$

$$\Rightarrow \ln L(\lambda) = n\ln\lambda - \lambda\sum_{i=1}^{n} x_i$$

$$\Rightarrow \frac{d}{d\lambda}\ln L(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Rightarrow \lambda = \frac{n}{\sum\limits_{i=1}^{n} x_i} = \frac{1}{\overline{x}} .$$

Vậy ước lượng hợp lý cực đại của tham số λ là $\Lambda = \dfrac{n}{\sum\limits_{i=1}^{n} X_i} = \dfrac{1}{\overline{X}}$.

5.25 $f = \dfrac{1082}{2000}$; Điều kiện $\begin{cases} nf = 1082 > 10 \\ n(1-f) = 918 > 10 \end{cases}$

$$f - U_{\alpha/2}\frac{\sqrt{f(1-f)}}{\sqrt{n}} = \frac{1082}{2000} - 2,33\frac{\sqrt{918\times 1082}}{2000\sqrt{2000}} = 0,515$$

Vậy tối thiểu có 51,5% số phiếu bầu cho ứng cử viên A.

5.26 $\overline{x} = \dfrac{\sum x_i}{n} = \dfrac{34,15}{35} = 0,976$.

$$s^2 = \frac{1}{n-1}\left[\sum {x_i}^2 - \frac{\left(\sum x_i\right)^2}{n}\right] = \frac{1}{34}\left[33,8943 - \frac{(34,15)^2}{35}\right] = 0,01687$$

$$\Rightarrow s = 0,1299;\ U_{\alpha/2}\frac{s}{\sqrt{n}} = 1,96\times\frac{0,1299}{\sqrt{35}} = 0,043.$$

Khoảng tin cậy 95%: $[0,933\ ;\ 1,019]$.

5.27 Tần suất mẫu $f = \dfrac{53}{400}$, điều kiện $\begin{cases} nf = 53 > 10 \\ n(1-f) = 347 > 10 \end{cases}$

Gọi p là xác suất bắt được con cá có đánh dấu, khoảng tin cậy 95% của p:

$$U_{\alpha/2}\frac{\sqrt{f(1-f)}}{\sqrt{n}} = 1,96\frac{\sqrt{53\times 347}}{400\sqrt{400}} = 0,0332$$

Khoảng ước lượng $[0,0993\ ;\ 0,1657]$

Mặt khác $p = \dfrac{2000}{N}$, trong đó N là số cá trong hồ.

Vậy $0,0993 < \dfrac{2000}{N} < 0,1657 \Rightarrow \dfrac{2000}{0,1657} < N < \dfrac{2000}{0,0993}$.

$$\Rightarrow 12070 < N < 20141$$

5.28 Đặt $u_i = \dfrac{x_i - 18,25}{5}$

$$\Rightarrow \overline{x} = 5\frac{\sum r_i u_i}{n} + 18,25 = 5\frac{-1,8}{40} + 18,25 = 18,025$$

$$s^2 = \frac{5^2}{n-1}\left[\sum r_i {u_i}^2 - \frac{\left(\sum r_i u_i\right)^2}{n}\right] = \frac{25}{39}\left[0,76 - \frac{(-1,8)^2}{40}\right] = 0,435$$

$$\Rightarrow s = 0,66;\ U_{\alpha/2}\frac{s}{\sqrt{n}} = 1,64 \times \frac{0,66}{\sqrt{40}} = 0,171$$

a) Khoảng tin cậy 90%: $[17,854;\ 18,196]$.

b) Kích thước mẫu cần thiết $n \geq \dfrac{U^2_{\alpha/2} s^2}{\varepsilon^2} = 116,99$ chọn $n = 117$

5.29 Đặt $u_i = x_i - 50 \Rightarrow \overline{x} = \dfrac{\sum r_i u_i}{n} + 50 = \dfrac{-8}{27} + 50 = 49,704$

$$U_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1,96 \times \frac{1}{\sqrt{27}} = 0,377$$

a) Khoảng tin cậy 95%: $[49,327\ ;\ 50,081]$.

b) Kích thước mẫu cần thiết: $n \geq \dfrac{U^2_{\alpha/2}\sigma^2}{\varepsilon^2} = 384,16$ chọn $n = 385$

5.30 Khoảng tin cậy 95% của phương sai được tính theo công thức (5.53).

$$\left(\frac{nS^2}{\chi^2_{\alpha/2}(n)}\ ;\frac{nS^2}{\chi^2_{1-\alpha/2}(n)}\right)$$

Tra bảng χ^2 với 15 bậc tự do và với giả thiết $\hat{S}^2 = 0,5$ ta tìm được khoảng tin cậy:

$$\left(\frac{25\cdot 0{,}5}{40{,}646};\frac{25\cdot 0{,}5}{13{,}120}\right)=(0{,}3075\ ;\ 0{,}9520).$$

5.31 Gọi μ là trọng lượng trung bình của một bao sản phẩm được đóng gói. Ta kiểm định giả thiết $H_0:\ \mu=100$; đối thiết $H_1:\ \mu<100$

Tiêu chuẩn kiểm định $\qquad T=\dfrac{(100-\bar{X})\sqrt{n}}{S}$;

Miền bác bỏ $\qquad W_\alpha=\{T>2{,}086\}.$

Đặt $u_i=\dfrac{x_i-99{,}25}{5}\ \Rightarrow\ \sum r_iu_i=0{,}4\ ;\ \sum r_iu_i^{\,2}=0{,}42$

$$\Rightarrow\ \bar{x}=5\times\frac{0{,}4}{29}+99{,}25=99{,}319\ ;$$

$$s^2=25\times\frac{1}{28}\left[0{,}42-\frac{0{,}4^2}{29}\right]=0{,}37\Rightarrow s=0{,}608$$

$$T_{qs}=\frac{(100-99{,}319)\sqrt{29}}{0{,}608}=6{,}032\in W_\alpha.$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là sản phẩm bị đóng thiếu.

5.32 Gọi μ là thời gian trung bình hoàn thành một sản phẩm.

Ta kiểm định giả thiết $H_0:\ \mu=14$; đối thiết $H_1:\ \mu\neq 14$

Tiêu chuẩn kiểm định: $T=\dfrac{(\bar{X}-14)\sqrt{n}}{S}$;

Miền bác bỏ: $W_\alpha=\{|T|>1{,}96\}.$

Đặt $u_i=\dfrac{x_i-15}{2}\ \Rightarrow\ \sum r_iu_i=0\ ;\ \sum r_iu_i^{\,2}=300$

$$\Rightarrow \bar{x} = 15\,; \quad s^2 = 4 \times \frac{1}{249}\left[300 - \frac{0}{300}\right] = 4{,}819 \Rightarrow s = 2{,}195$$

$$\Rightarrow \quad T_{qs} = \frac{(115-14)\sqrt{300}}{2{,}195} = 7{,}89 \in W_\alpha\,.$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là cần thay đổi định mức.

5.33 Gọi μ là mức hao phí xăng trung bình của ôtô chạy từ A đến B. Ta kiểm định giả thiết $H_0: \mu = 50$; đối thiết $H_1: \mu < 50$

Tiêu chuẩn kiểm định $\qquad T = \dfrac{\left(50 - \bar{X}\right)\sqrt{n}}{S}$;

Miền bác bỏ $\qquad W_\alpha = \{T > 2{,}052\}$.

Theo mẫu ta có $\quad \bar{x} = \dfrac{1387{,}5}{28} = 49{,}5536$;

$$s^2 = \frac{1}{27}\left(6876375 - \frac{1387{,}5^2}{28}\right) = \frac{8{,}1696}{27} = 0{,}3026 \;\Rightarrow s = 0{,}55$$

$$T_{qs} = \frac{(50 - 49{,}53)\sqrt{30}}{0{,}55} = 4{,}2948 \in W_\alpha\,.$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là mức hao phí xăng có giảm xuống.

5.34 Gọi μ là số hoá đơn trung bình hệ thống máy tính mới xử lý được trong 1 giờ. Ta kiểm định giả thiết: $H_0: \mu = 1300$; đối thiết: $H_1: \mu > 1300$

Tiêu chuẩn kiểm định $\qquad T = \dfrac{\left(\bar{X} - 1300\right)\sqrt{n}}{S}$;

Miền bác bỏ $\qquad W_\alpha = \{T > 1{,}96\}$.

Từ mẫu cụ thể ta có

$$T = \frac{(1378-1300)\sqrt{40}}{215} = 2,294 > 1,96$$

Vậy bác bỏ H_0 chấp nhận H_1, nghĩa là hệ thống máy tính mới xử lý tốt hơn.

Đáp án chương 6

Câu hỏi	6.1	6.2	6.3	6.4	6.5	6.6
Đáp án	Sai	Sai	Đúng	Đúng	Đúng	Sai

6.7 $P\{X_0 = 0, X_1 = 2, X_2 = 1\}$

$= P\{X_0 = 0\} P\{X_1 = 2 | X_0 = 0\} P\{X_2 = 1 | X_0 = 0, X_1 = 2\}$

$= P\{X_0 = 0\} P\{X_1 = 2 | X_0 = 0\} P\{X_2 = 1 | X_1 = 2\}$

$= 0,3 \cdot 0,7 \cdot 0,8 = 0,168.$

6.8 Điều kiện thứ hai của ma trận Markov có thể viết dưới dạng

$$\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \text{ hay là } P\mathbf{a} = \mathbf{a} \text{ trong đó } \mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Do đó $P^2\mathbf{a} = P\mathbf{a} = \mathbf{a} \ \ldots \Rightarrow P^n\mathbf{a} = P^{n-1}\mathbf{a} = \mathbf{a}$.

6.9 a) $P^2 = \begin{bmatrix} 0,47 & 0,13 & 0,40 \\ 0,42 & 0,14 & 0,44 \\ 0,26 & 0,17 & 0,57 \end{bmatrix}.$

b) $P\{X_3 = 1 | X_1 = 0\} = P\{X_2 = 1 | X_0 = 0\} = 0,13;$

$$P\{X_3=1|X_0=0\}=P\{X_3=1,X_2=0|X_0=0\}$$
$$+P\{X_3=1,X_2=1|X_0=0\}$$
$$+P\{X_3=1,X_2=2|X_0=0\}$$
$$=P\{X_2=0|X_0=0\}P\{X_3=1|X_0=0,X_2=0\}$$
$$+P\{X_2=1|X_0=0\}P\{X_3=1|X_0=0,X_2=1\}$$
$$+P\{X_2=2|X_0=0\}P\{X_3=1|X_0=0,X_2=2\}$$
$$=P\{X_2=0|X_0=0\}P\{X_3=1|X_2=0\}$$
$$+P\{X_2=1|X_0=0\}P\{X_3=1|X_2=1\}$$
$$+P\{X_2=2|X_0=0\}P\{X_3=1|X_2=2\}$$
$$=0,47\cdot 0,2+0,13\cdot 0,2+0,40\cdot 0,1=0,16.$$

c) Phân bố dừng $[x,y,z]$ là nghiệm của hệ phương trình

$$\begin{cases}[x \quad y \quad z]P=[x \quad y \quad z]\\ x,y,z\geq 0;\ x+y+z=1\end{cases}.$$

Như vậy x,y,z là nghiệm không âm của hệ phương trình

$$\begin{cases}-9x+2y+6z=0\\ 2x-8y+z=0\\ x+y+z=1\end{cases}$$

Hệ phương trình có nghiệm $x=\dfrac{50}{139}$, $y=\dfrac{21}{139}$, $z=\dfrac{68}{139}$.

6.10 Đặt $p_0=P\{X_0=0\}$.

a) $P\{X_0=0,X_1=0,X_2=0\}=P\{X_0=0\}P\{X_1=0,X_2=0|X_0=0\}$

$=P\{X_0=0\}P\{X_1=0|X_0=0\}P\{X_2=0|X_0=0,X_1=0\}=p_0(1-\alpha)^2$.

b) $P\{X_0=0,X_1=0,X_2=0\}+P\{X_0=0,X_1=1,X_2=0\}$
$= p_0\left(\alpha^2+(1-\alpha)^2\right)$.

c) $P\{X_5=0|X_0=0\}=16\alpha^5+40\alpha^4+40\alpha^3+20\alpha^2+5\alpha+1$.

6.11 a) $P=(0,2)(0,3)(0,2)(0,5)(0,2)=0,0012$.

b) $P^2=\begin{bmatrix}0,21 & 0,26 & 0,26 & 0,27\\ 0,20 & 0,28 & 0,24 & 0,28\\ 0,18 & 0,30 & 0,21 & 0,31\\ 0,22 & 0,32 & 0,22 & 0,24\end{bmatrix}$

$P\{X_1=a,X_3=c,X_4=b,X_5=a|X_0=b\}$

$=P\{X_1=a|X_0=b\}P\{X_3=c|X_1=a\}P\{X_4=b|X_3=c\}P\{X_5=a|X_4=b\}$

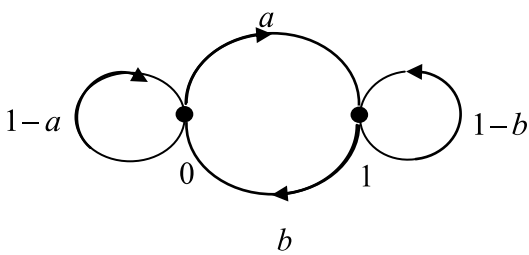
$=(0,2)(0,26)(0,5)(0,2)=0,0052$.

c) $P^5=\begin{bmatrix}0,2029 & 0,2912 & 0,2319 & 0,2740\\ 0,2028 & 0,2914 & 0,2317 & 0,2740\\ 0,2027 & 0,2918 & 0,2313 & 0,2742\\ 0,2030 & 0,2916 & 0,2317 & 0,2737\end{bmatrix}$

d) $P\{X_5=a|X_0=b\}=0,2028$

e) $\mathbf{P}(5)=[0,2028 \quad 0,2915 \quad 0,2316 \quad 0,2740]$.

6.12 a)



b) Hệ phương trình:

$$\begin{cases} \begin{bmatrix} 1-a & b \\ a & 1-b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix} \\ x \geq 0, y \geq 0; x+y=1 \end{cases} \Leftrightarrow \begin{cases} ax - by = 0 \\ x + y = 1 \\ x \geq 0, y \geq 0 \end{cases} \Leftrightarrow \begin{cases} x = \dfrac{b}{a+b} \\ y = \dfrac{a}{a+b} \end{cases}$$

c) $P^n = \dfrac{1}{a+b}\left\{ \begin{bmatrix} b & a \\ b & a \end{bmatrix} + (1-a-b)^n \begin{bmatrix} a & -a \\ -b & b \end{bmatrix} \right\}$

$$\mathbf{P}(n) = \mathbf{P}(0)P^n = \begin{bmatrix} 0,5 & 0,5 \end{bmatrix} \frac{1}{a+b}\left\{ \begin{bmatrix} b & a \\ b & a \end{bmatrix} + (1-a-b)^n \begin{bmatrix} a & -a \\ -b & b \end{bmatrix} \right\}$$

$$= \frac{1}{2(a+b)} \begin{bmatrix} 2b + (1-a-b)^n (a-b) & 2a + (1-a-b)^n (b-a) \end{bmatrix}$$

d) Nếu $a+b \leq 1$ thì $\lim\limits_{n\to\infty} P^n = \dfrac{1}{a+b} \begin{bmatrix} b & a \\ b & a \end{bmatrix}$

6.13 Không gian trạng thái sẽ là $E = \{-1, 0, 1, 2, 3\}$.

Theo công thức (6.21) ta có:

$$p_{ij} = P\{X(n+1) = j \,|\, X(n) = i\} = \begin{cases} P\{\xi = 3 - j\} & \text{nếu} \quad i \leq 0, \\ P\{\xi = i - j\} & \text{nếu} \quad 0 < i \leq 3. \end{cases}$$

$$p_{-1,-1} = P\{X(n+1) = -1 \,|\, X(n) = -1\} = P(\phi) = 0,$$

$$p_{-1,0} = P\{X(n+1) = 0 \,|\, X(n) = -1\} = P(\xi = 3) = P(\phi) = 0,$$

$$p_{-1,1} = P\{X(n+1) = 1 \,|\, X(n) = -1\} = P(\xi = 2) = 0,3,$$

$$p_{-1,2} = P\{X(n+1) = 2 \,|\, X(n) = -1\} = P(\xi = 1) = 0,3,$$

$$p_{-1,3} = P\{X(n+1) = 1 \,|\, X(n) = -1\} = P(\xi = 0) = 0,4,$$

. . . . . .

Ma trận xác suất chuyển:

$$P=\begin{bmatrix} 0 & 0 & 0,3 & 0,3 & 0,4 \\ 0 & 0 & 0,3 & 0,3 & 0,4 \\ 0,3 & 0,3 & 0,4 & 0 & 0 \\ 0 & 0,3 & 0,3 & 0,4 & 0 \\ 0 & 0 & 0,3 & 0,3 & 0,4 \end{bmatrix}$$

6.14 *A* có 43,3% và *B* có 56,7% thị phần.

6.15 a) $P=\begin{bmatrix} 3/5 & 1/5 & 1/5 \\ 1/6 & 2/3 & 1/6 \\ 3/8 & 3/8 & 1/4 \end{bmatrix}$

$$\mathbf{P}(1)=\begin{bmatrix} 2/10 & 3/10 & 5/10 \end{bmatrix}\begin{bmatrix} 3/5 & 1/5 & 1/5 \\ 1/6 & 2/3 & 1/6 \\ 3/8 & 3/8 & 1/4 \end{bmatrix}$$

$$=\begin{bmatrix} 143/400 & 171/400 & 86/400 \end{bmatrix}$$

b) $P^n=\begin{bmatrix} 0,3659 & 0,4390 & 0,1951 \\ 0,3659 & 0,4390 & 0,1951 \\ 0,3659 & 0,4390 & 0,1951 \end{bmatrix}$ với mọi $n\geq 15$

Vậy chuỗi có phân bố giới hạn $\mathbf{P}^*=\begin{bmatrix} 0,3659 & 0,4390 & 0,1951 \end{bmatrix}$.

6.16 1) $i\leftrightarrow i$ vì $p_{ii}^{(0)}=1$

2) $i\leftrightarrow j$ thì $j\leftrightarrow i$

3) Để chứng minh tính bắc cầu ta chỉ cần chứng minh:

$i\to j$ và $j\to k$ thì $i\to k$.

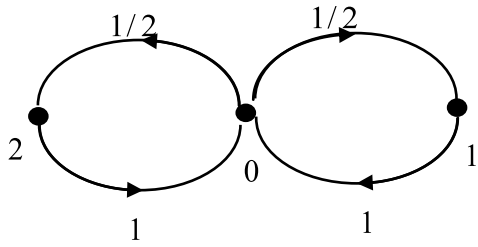
Nếu $i\to j$ và $j\to k$ tthì tồn tại n,m sao cho $p_{ij}^{(n)}>0$ và $p_{jk}^{(m)}>0$.

Theo phương trình Chapman-Kolmogorov ta có

$$p_{ik}^{(n+m)} = \sum_r p_{ir}^{(n)} p_{rk}^{(m)} \geq p_{ij}^{(n)} p_{jk}^{(m)} > 0.$$

Do đó $i \to k$.

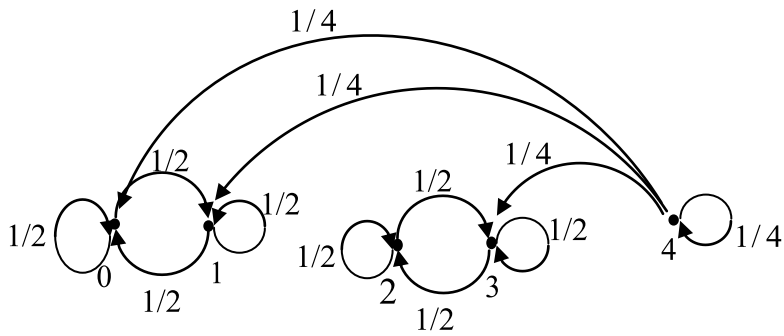
6.17 a)



b) $P^{(2)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}$,

$$P^{(3)} = P \Rightarrow P^{(2n)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix},\ P^{(2n+1)} = P.$$

6.18 Các trạng thái có chu kỳ 1. Có 3 lớp liên thông là $\{0,1\}$, $\{2,3\}$, $\{4\}$.



6.19 Áp dụng công thức (6.59) với $p = 0,4$ ta có phân bố dừng

$$p_0 = \frac{1}{6}, p_1 = \frac{5}{18}, ..., p_j = \frac{5}{18}\left(\frac{2}{3}\right)^{j-1}; j \geq 2.$$

6.20 Áp dụng công thức (6.61) với $p = 0,6$, $N = 10$ ta có phân bố dừng

$$p_i = \frac{p_0}{q}\left(\frac{p}{q}\right)^{i-1}, 1 \leq i \leq N-1; \; p_N = p_0\left(\frac{p}{q}\right)^{N-1}$$

$$p_0 = \frac{1}{1+(p/q)^{N-1} + \frac{1}{q}\sum_{j=1}^{N-1}(p/q)^{j-1}}$$

$$p_0 = 0,004; \; p_1 = \frac{p_0}{q} = 0,01; \; p_2 = \frac{p_0}{q}(p/q) = 0,015;$$

$$p_3 = \frac{p_0}{q}(p/q)^2 = 0,0225 \; \ldots$$

$$p_9 = \frac{p_0}{q}(p/q)^8 = 0,2563; \; p_{10} = p_0(p/q)^9 = 0,1538.$$

PHỤ LỤC

Phụ lục 1: GIÁ TRỊ HÀM MẬT ĐỘ XÁC SUẤT PHÂN BỐ CHUẨN TẮC

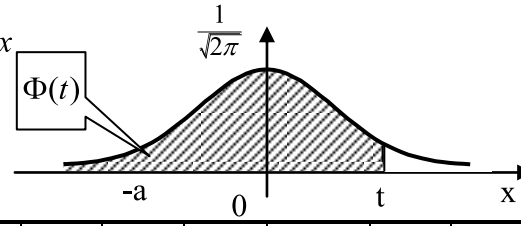
$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2370	2347	2320	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551

2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	000065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	00080	0008	0007	0007	0007	0007	0006
3,6	0005	0005	0005	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

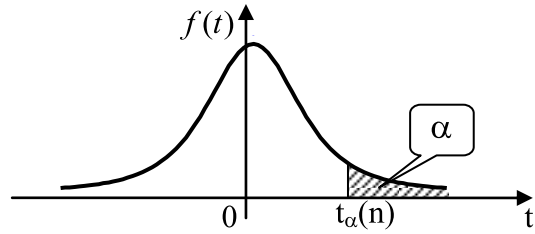
Phụ lục 2: GIÁ TRỊ HÀM PHÂN BỐ CHUẨN TẮC

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$



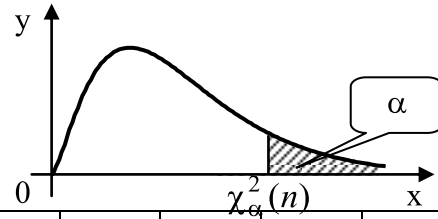
t	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0,1	5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0,2	5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0,3	6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0,4	6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0,5	0,6915	6950	6985	7019	7054	7088	7123	7156	7190	7224
0,6	7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0,7	7580	7611	7642	7673	7703	7734	7764	7794	7823	7852
0,8	7881	7910	7939	7967	7995	8023	8051	8078	8106	8132
0,9	8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1,0	0,8413	8438	8461	8485	8508	8531	8554	8577	8599	8621
1,1	8643	8665	8686	8708	8729	8749	8770	8790	8810	8830
1,2	8849	8869	8888	8907	8925	8944	8962	8980	8997	9015
1,3	9032	9049	9066	9082	9099	9115	9131	9147	9162	9177
1,4	9192	9207	9222	9236	9251	9265	9279	9292	9306	9319
1,5	0,9332	9345	9357	9370	9382	9394	9406	9418	9429	9441
1,6	9452	9463	9474	9484	9495	9505	9515	9525	9535	9545
1,7	9554	9564	9573	9582	9591	9599	9608	9616	9625	9633
1,8	9641	9649	9656	9664	9671	9678	9686	9693	9699	9706
1,9	9712	9719	9726	9732	9738	9744	9750	9756	9761	9767
2,0	0,9773	9778	9783	9788	9793	9798	9803	9808	9812	9817
2,1	9821	9826	9830	9834	9838	9842	9846	9850	9854	9857
2,2	9861	9864	9868	9871	9875	9878	9881	9884	9887	9890
2,3	9893	9896	9898	9901	9904	9906	9909	9911	9913	9916
2,4	9918	9920	9922	9925	9927	9929	9931	9932	9934	9936
2,5	0,9938	9940	9941	9943	9945	9946	9948	9949	9951	9952
2,6	9953	9955	9956	9957	9959	9960	9961	9962	9963	9964
2,7	9965	9966	9967	9968	9969	9970	9971	9972	9973	9974
2,8	9974	9975	9976	9977	9977	9978	9979	9979	9980	9981
2,9	9981	9982	9982	9983	9984	9984	9985	9985	9986	9986
t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9
$\Phi(t)$	0,9987	9990	9993	9995	9996	9997	9998	9999	9999	9999

Phụ lục 3: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ STUDENT



Bậc tự do	$\alpha = 0,05$	$\alpha = 0,025$	$\alpha = 0,01$	$\alpha = 0,005$	$\alpha = 0,002$
1	6,314	12,706	31,821	63,657	318,309
2	2,920	4,303	6,965	9,925	22,327
3	2,353	3,128	4,541	5,841	10,215
4	2,132	2,776	3,747	4,604	7,173
5	2,015	2,571	3,365	4,032	5,893
6	1,943	2,447	3,143	3,707	5,208
7	1,895	2,365	2,998	3,499	4,705
8	1,860	2,306	2,896	3,355	4,501
9	1,833	2,262	2,821	3,250	4,297
10	1,812	2,228	2,764	3,169	4,144
11	1,796	2,201	2,718	3,106	4,025
12	1,782	2,179	2,681	3,055	3,930
13	1,771	2,160	2,650	3,012	3,852
14	1,761	2,145	2,624	2,977	3,787
15	1,753	2,131	2,606	2,947	3,733
16	1,746	2,120	2,583	2,921	3,686
17	1,740	2,110	2,567	2,898	3,646
18	1,734	2,101	2,552	2,878	3,610
19	1,729	2,093	2,539	2,861	3,579
20	1,725	2,086	2,528	2,845	3,552
21	1,721	2,080	2,518	2,831	3,527
22	1,717	2,074	2,508	2,819	3,505
23	1,714	2,069	2,500	2,807	3,485
24	1,711	2,064	2,492	2,797	3,467
25	1,708	2,060	2,485	2,787	3,450
26	1,796	2,056	2,479	2,779	3,435
27	1,703	2,052	2,473	2,771	3,421
28	1,701	2,048	2,467	2,763	3,408
29	1,699	2,045	2,462	2,756	3,396
inf	1,645	1,960	2,326	2,576	3,090

Phụ lục 4: GIÁ TRỊ TỚI HẠN CỦA PHÂN BỐ "KHI BÌNH PHƯƠNG" χ^2



Bậc tự do	$\chi^2_{0,995}$	$\chi^2_{0,99}$	$\chi^2_{0,97}$	$\chi^2_{0,95}$	$\chi^2_{0,05}$	$\chi^2_{0,025}$	$\chi^2_{0,01}$	$\chi^2_{0,005}$
1	0,000	0,000	0,001	0,004	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	11,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,982	22,362	24,736	27,688	28,819
14	4,075	4,660	5,629	6,571	23,685	26,119	29,141	31,319
15	5,001	5,229	6,262	7,261	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	28,869	31,524	34,805	37,156
19	6,844	7,633	8,907	10,117	30,144	32,852	36,191	38,582
20	7,343	8,260	9,591	10,851	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	32,671	35,479	38,932	41,401
22	8,543	9,542	10,982	12,388	33,924	36,781	30,289	42,796
23	9,260	10,196	11,689	13,091	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	36,415	39,364	42,980	45,558
25	10,520	11,524	13,120	14,611	37,625	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	40,113	43,194	46,993	46,645
28	12,461	13,565	15,308	16,928	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	42,557	45,722	49,588	52,336
30	13,787	14,930	16,791	18,493	43,773	46,979	50,892	53,672

Phụ lục 5: GIÁ TRỊ HÀM KHỐI LƯỢNG XÁC SUẤT POISSON

$$P\{X = k\} = \frac{e^{-\lambda}\lambda^k}{k!}$$

k \ λ	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	0,090484	0,163746	0,222245	0,268128	0,303265	0,329287
2	0,004524	0,016375	0,033337	0,053626	0,075817	0,098786
3	0,000151	0,001091	0,003334	0,007150	0,012636	0,019757
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964
5		0,000002	0,000015	0,000057	0,000158	0,000356
6			0,000001	0,000004	0,000013	0,000035
7					0,000001	0,000003

k \ λ	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367877	0,135335	0,049787
1	0,347610	0,359463	0,365913	0,367789	0,270671	0,149361
2	0,121663	0,143785	0,164661	0,183940	0,270671	0,224042
3	0,028388	0,038343	0,049398	0,061313	0,180447	0,224042
4	0,004968	0,007669	0,011115	0,015328	0,090224	0,168031
5	0,000695	0,001227	0,002001	0,003066	0,036089	0,100819
6	0,000081	0,000164	0,000300	0,000511	0,012030	0,050409
7	0,000008	0,000019	0,000039	0,000073	0,003437	0,021604
8		0,000002	0,000004	0,000009	0,000859	0,008101
9				0,000001	0,000191	0,002701
10					0,000038	0,000810
11					0,000007	0,000221
12					0,000001	0,000055
13						0,000013
14						0,000003
15						0,000001

k \ λ	4,0	5,0	6,0	7,0	8,0	9,0
0	0,018316	0,006738	0,002479	0,000912	0,000335	0,000123
1	0,073263	0,033690	0,014873	0,006383	0,002684	0,001111
2	0,146525	0,084224	0,044618	0,022341	0,010735	0,004998
3	0,195367	0,140374	0,089235	0,052129	0,028626	0,014994
4	0,195367	0,175467	0,133853	0,191226	0,057252	0,033737
5	0,156293	0,175467	0,160623	0,127717	0,091604	0,060727
6	0,104194	0,146223	0,160623	0,149003	0,122138	0,091090
7	0,059540	0,104445	0,137677	0,149003	0,139587	0,117116
8	0,029770	0,065278	0,103258	0,130377	0,139587	0,131756
9	0,013231	0,036266	0,068838	0,011405	0,120477	0,131756
10	0,005292	0,018133	0,041303	0,070983	0,099262	0,118580
11	0,001925	0,008242	0,022529	0,045171	0,072190	0,097020
12	0,000642	0,003434	0,011262	0,026350	0,048127	0,072765
13	0,000197	0,001321	0,005199	0,014188	0,029616	0,050376
14	0,000056	0,000472	0,002228	0,007094	0,013924	0,032384
15	0,000015	0,000157	0,000891	0,003311	0,009026	0,019431
16	0,000004	0,000049	0,000334	0,001448	0,004513	0,010930
17	0,000001	0,000014	0,000118	0,000596	0,002124	0,005786
18		0,000004	0,000039	0,000232	0,000944	0,002893
19		0,000001	0,000012	0,000085	0,000397	0,001370
20			0,000004	0,000030	0,000159	0,000617
21			0,000001	0,000010	0,000061	0,000264
22				0,000003	0,000022	0,000108
23				0,000001	0,000008	0,000042
24					0,000003	0,000016
25					0,000001	0,000006
26						0,000002
27						0,000001

Phụ lục 6: GIÁ TRỊ HÀM PHÂN BỐ POISSON

$$P\{X \le k\} = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^{i}}{i!}$$

k \ λ	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670320	0,606531	0,548812
1	0,995321	0,982477	0,963063	0,938448	0,909796	0,878099
2	0,999845	0,998853	0,996400	0,992074	0,985612	0,976885
3	0,999996	0,999943	0,999734	0,999224	0,998248	0,996642
4	1,000000	0,999998	0,999984	0,999939	0,999828	0,999606
5		1,000000	0,999999	0,999996	0,999986	0,999962
6				1,000000	0,999999	0,999997
7						1,000000

k \ λ	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496585	0,449329	0,406570	0,367877	0,135335	0,049787
1	0,844195	0,808792	0,772483	0,735759	0,406006	0,199148
2	0,965858	0,952577	0,937144	0,919699	0,676677	0,423190
3	0,994246	0,990920	0,986542	0,981012	0,857124	0,647232
4	0,999214	0,998589	0,997657	0,996340	0,947348	0,815263
5	0,999909	0,999816	0,999658	0,999403	0,983437	0,916082
6	0,999990	0,999980	0,999958	0,999917	0,995467	0,966491
7	0,999999	0,999998	0,999997	0,999990	0,998904	0,988095
8			1,000000	0,999999	0,999753	0,996196
9				1,000000	0,999954	0,998897
10					0,999992	0,999707
11					0,999999	0,999928
12					1,000000	0,999983
13						0,999996
14						0,999999
15						1,000000

k \ λ	4,0	5,0	6,0	7,0	8,0	9,0
0	0,018316	0,006738	0,002479	0,000912	0,000335	0,000123
1	0,091579	0,040428	0,017352	0,007295	0,003019	0,001234
2	0,238105	0,124652	0,061970	0,029636	0,013754	0,006232
3	0,433472	0,265026	0,151205	0,081765	0,042380	0,021228
4	0,785132	0,615960	0,445681	0,300708	0,191236	0,115690
5	0,889326	0,762183	0,606304	0,449711	0,313374	0,206780
6	0,948866	0,866628	0,743981	0,598711	0,452961	0,323896
7	0,978636	0,931806	0,847239	0,729091	0,592548	0,455652
8	0,991867	0,968172	0,916077	0,830496	0,716625	0,587408
9	0,997159	0,986305	0,957380	0,901479	0,815887	0,705988
10	0,999084	0,984547	0,979909	0,946650	0,888077	0,803008
11	0,999726	0,997981	0,991173	0,973000	0,936204	0,875773
12	0,999923	0,999202	0,996372	0,987188	0,965820	0,926149
13	0,999979	0,999774	0,998600	0,994282	0,982744	0,958533
14	0,999994	0,999931	0,999491	0,997593	0,991770	0,977964
15	0,999998	0,999980	0,999825	0,999041	0,996283	0,988894
16	0,999999	0,999994	0,999943	0,999637	0,998407	0,994680
17	0,999999	0,999998	0,999982	0,999869	0,999351	0,997573
18	0,999999	0,999999	0,999994	0,999955	0,999748	0,998943
19	1,000000	0,999999	0,999998	0,999985	0,999907	0,999560
20		1,000000	0,999999	0,999995	0,999967	0,999824
21			0,999999	0,999998	0,999989	0,999932
22			1,000000	0,999999	0,999997	0,999974
23				0,999999	0,999998	0,999990
24				1,000000	0,999999	0,999996
25					1,000000	0,999998
26						0,999999
27						1,000000

BẢNG CHỈ DẪN THUẬT NGỮ

Bảng phân bố xác suất biên 119
Bảng phân bố ghép lớp 187
Bảng phân bố tần số thực nghiệm 191
Bảng phân bố tần suất thực nghiệm 191
Bất đẳng thức Markov 173
Bất đẳng thức Trêbưsép 174
Biểu đồ tần số hình gậy 194
Biểu đồ đa giác tần suất 194
Biến cố sơ cấp 13
Biến cố 12
Biến cố chắc chắn 13
Biến cố không thể 14
Biến cố đối 12
Biến cố xung khắc 15
Biến cố độc lập 36
Biến ngẫu nhiên 119,196
Biến ngẫu nhiên kỳ vọng có điều kiện 159
Biểu đồ chuyển trạng thái 268
Cá thể 168
Các trạng thái liên thông 266
Chỉnh hợp 21
Chu kỳ của trạng thái 267
Chuỗi Markov thời gian rời rạc thuần nhất 249
Phân bố dừng 258
Phân bố giới hạn 258
Phân bố ergodic 258
Phương sai 56,129
Phân vị 97
Phân bố có điều kiện 153
Phân bố chuẩn n chiều 160
Phương pháp hợp lý cực đại 206
Phương pháp moment 210
Phương sai mẫu 197
Phương trình Chapman-Kolmogorov 251
Quá trình ngẫu nhiên 189,241
Quá trình độc lập 245
Quá trình Bernoulli 245
Quá trình gia số độc lập 246
Quá trình gia số độc lập dừng 246
Quá trình Martingal 246
Quá trình Markov 247
Quá trình dừng 248
Quy tắc hai xích ma và ba xích ma 86
Sai lầm loại một và sai lầm loại hai 225
Tần suất mẫu 232
Tính độc lập của các biến ngẫu nhiên 127

Công thức xác suất đầy đủ 40
Công thức Bayes 41
Dấu hiệu nghiên cứu 198
Di động ngẫu nhiên trên đường thẳng 276
Định lý giới hạn trung tâm 179
Độ chính xác của ước lượng 213
Độ lệch chuẩn mẫu 215
Giá trị quan sát của tiêu chuẩn kiểm định 225
Giả thiết thống kê 172
Hàm phân bố xác suất 58,178
Hàm khối lượng xác suất 64
Hàm mật độ xác suất 75
Hệ số bất đối xứng 101
Hàm khối lượng xác suất đồng thời 118
Hàm mật độ xác suất đồng thời 122
Hàm mật độ xác suất biên 123
Hàm đặc trưng 102
Hàm phân bố xác suất đồng thời 169
Hàm của một biến ngẫu nhiên 135
Hàm của hai biến ngẫu nhiên 139
Hàm hợp lý 207
Hàm phân bố thực nghiệm của mẫu 192
Hàm mẫu của quá trình ngẫu nhiên 142
Hệ số nhọn 101
Hiệp phương sai 113
Hệ đầy đủ biến cố 15
Thủ tục kiểm định giả thiết thống kê 226
Tích biến cố 15
Tổ hợp 12
Tổ chức đồ 187
Thủ tục kiểm định giả thiết thống kê 226
Tổng thể 188,189
Tổng của hai biến cố 14
Trạng thái tuần hoàn 267
Trạng thái hồi quy 271
Trạng thái không hồi quy 271
Trạng thái hồi quy dương 273
Trung bình mẫu 197
Trung vị 97
Ước lượng điểm 203
Ước lượng không chệch 204
Ước lượng hiệu quả 204
Ước lượng vững 205
Véc tơ ngẫu nhiên 114
Xác suất có điều kiện 33
Phân bố dừng 258
Phân bố giới hạn 258
Phân bố ergodic 258
Phương sai 56,129
Phân vị 97
Phân bố có điều kiện 153
Phân bố chuẩn nhiều chiều 160

Hoán vị 21
Hội tụ theo xác suất 205
Hội tụ theo phân bố 171
Khoảng tin cậy 188
Không gian mẫu 13
Không gian trạng thái 241
Kỳ vọng 56,113
Kỳ vọng có điều kiện 153
Luật số lớn Trêbưsép 175
Luật số lớn Bernoulli 177
Ma trận hiệp phương sai 131
Ma trận xác suất chuyển 249
Mẫu ngẫu nhiên 114
Miền bác bỏ 224
Phép thử Bernoulli 46
Phân bố Bernoulli 68
Phân bố nhị thức 69
Phân bố Poission 72
Phân bố đều 78
Phân bố mũ 79
Phân bố Erlang 81
Phân bố chuẩn 81
Phân bố “khi bình phương” 87
Phân bố Student 88
Phân bố xác suất của chuỗi tại thời điểm thứ n 252
Mốt 98
Moment 101
Mức ý nghĩa của kiểm định 224
Phép thử 24

TÀI LIỆU THAM KHẢO

[1]. Nguyễn Phạm Anh Dũng, 1999*; Các hàm và xác suất ứng dụng trong viễn thông,* Trung tâm Đào tạo Bưu chính Viễn thông 1.

[2]. Tống Đình Quỳ, *Hướng dẫn giải bài tập xác suất thống kê*, NXB Đại học Quốc gia Hà Nội, 2004.

[3]. Đặng Hùng Thắng, 1997; *Mở đầu về lý thuyết xác suất và các ứng dụng*, NXB Giáo dục.

[4]. Đặng Hùng Thắng; *Bài tập xác suất,* NXB Giáo dục, 1998.

[5]. Đặng Hùng Thắng; *Thống kê và ứng dụng,* NXB Giáo dục, 1999.

[6]. Nguyễn Duy Tiến, Vũ Việt Yên, 2000; *Lý thuyết xác suất*, NXB Giáo dục.

[7]. Nguyễn Duy Tiến (và tập thể), 2000. *Các mô hình xác suất và ứng dụng*, tập 1, 2, 3, NXB Đại học Quốc gia Hà Nội.

[8]. Trần Mạnh Tuấn; *Xác suất và Thống kê, lý thuyết và thực hành tính toán,* NXB Đại học Quốc gia Hà Nội, 2004.

[9]. Nguyễn Cao Văn và Trần Thái Ninh, *Bài giảng xác suất và thống kê toán,* NXB Thống kê, Hà Nội 1999.

[10]. Nguyễn Cao Văn, Trần Thái Ninh và Nguyễn Thế Hệ, *Bài tập lý thuyết xác suất và thống kê toán,* NXB Giáo dục, Hà Nội 2002.

[11]. J. L. Doob, 1953. *Stochastic Processes*; Willey and Sons, New York.

[12]. B.V. Gnedenko, *The theory of probability,* Mir publishers, Moscow 1976.

[13]. HWEI P. HSU, Ph.D., *Theory and problems Probability, Random variables, and Random processes,* Schaum's outline series. McGRAW=HILL, 1996.

[14]. S. Karlin, 1966. *A first Course in Stochastic Processes*. Academic Press, New York and London.

[15]. M. Loeve, 1977. *Probability Theory, I, II. 4th ed*, Springer - Verlag, Berlin and New York.

[16]. D. L. (Paul) Minh, *Applied Probability Models*, Duxbury, Thomson Learning, 2001.